Award Number: W81XWH-11-2-0158

TITLE: "Medical Robotic and Tele-surgical Simulation Education Research"

PRINCIPAL INVESTIGATOR: Roger D. Smith, PhD

CONTRACTING ORGANIZATION: Adventist Health System/Sunbelt, Inc.
Orlando, FL 32803

REPORT DATE: May 2017

TYPE OF REPORT: FINAL

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
May 2017

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
01 Sept 2011 - 28 Feb 2017

4. TITLE AND SUBTITLE
"Medical Robotic and Tele-surgical Simulation Education Research"

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-11-2-0158

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Roger D. Smith, PhD

Email: roger.smith@flhosp.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Adventist Health System/Sunbelt, Inc.

601 E Rollins Street

Orlando, Florida 32803-1248

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command

Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Purpose: Advance knowledge of robotic telesurgery and applications for surgical simulation.
Scope: Explore applications of surgical robots to telesurgery, evaluate effectiveness of simulator devices, and develop designs for future simulator devices.
Major Findings: This research agreement resulted in the creation of multiple educational products or prototypes in the field of robotic surgery. Prototypes have been carried to product completion through additional investment by Adventist Health System/Sunbelt, Inc. dba Florida Hospital and are available either freely or for purchase on the internet. The experiments have also demonstrated the effectiveness of multiple simulator devices in the field of robotic surgery and measuring the skills of multiple populations with these devices. Finally, we have demonstrated that remote telesurgery between hospital systems in the USA which are equipped with modern IT infrastructures are currently capable of supporting safe telesurgery using surgical robots.

15. SUBJECT TERMS
Telesurgery, robotic surgery, surgical simulation, simulator effectiveness, simulator design

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	155	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Table of Contents   Page

Introduction

This research agreement has resulted in the creation of multiple educational products or prototypes in the field of robotic surgery. Prototypes have been carried to product completion through additional investment by Adventist Health System/Sunbelt, Inc. dba Florida Hospital and are available either freely or for purchase on the internet. The experiments have also demonstrated the effectiveness of multiple simulator devices in the field of robotic surgery and measuring the skills of multiple populations with these devices. Finally, we have demonstrated that remote telesurgery between hospital systems in the USA which are equipped with modern IT infrastructures are currently capable of supporting safe telesurgery using surgical robots.

Overview

This cooperative agreement spanned a period of six years in two distinct phases:

Phase 1: Original Award, September 2011 – August 2014.

Phase 2: Awarded Extension, September 2014 – February 2017

Each phase carried a different, but related SOW, both of which are included in this report for reference. Within each phase there were multiple tasks and experiments. Each of these is summarized in this report.

This project was broken into three focus areas: robotic curriculum, simulation, and telesurgery. In each we explored various applications and extensions of the existing robotic surgical systems. Under robotic curriculum, we developed and validated a standardized curriculum for teaching robotic surgical techniques to surgeons. This curriculum and the supporting products we prototyped, came from the minds of 80 of the leading robotic surgeons in the world. Under simulation, we conducted multiple experiments into the capabilities and designs for simulators of robotic surgical devices. The results of each of these experiments has been published in a journal or presented at a conference. Under telesurgery, we measured the latency levels which could be tolerated by human surgeons using robotic systems in a remote telesurgery environment. We also measured the latency of robotic data through established hospital networks to determine whether current network performance could deliver data with latency below interference levels for human surgeons.

Statements of Work

Phase 1: Original Award SOW

There are three primary areas of this research: Telesurgery, Simulation, and Robotic Curriculum. (1) The telesurgery project will identify the characteristics of latency during telesurgery and investigate the application of principles of automatic surgery. (2) Under simulation, we will validate a simulator that can be used by military surgeons to maintain their robotic skills while deployed. We will then use this device to explore the feasibility of surgical rehearsal as a potential solution to the latency issue in telesurgery. (3) We will organize robotic surgery experts to develop a nationally accepted curriculum in the Fundamentals of Robotic Surgery (FRS).

Period 1

Telesurgery: Communications Latency Experiments. Identify communication latency, measure safe latency levels for each robotic movement, modify surgical procedures to be effective in this environment.

Milestone: Telesurgery latency experiment report. Award + 270 days

Simulation: Military-use Validation. Validate a robotic simulator for maintaining the robotic surgery skills of deployed military surgeons.

Milestone: Robotic simulator validation report. Award + 210 days

Robotic Curriculum: Consensus Conferences. Organize and host conferences of approximately 40 leading robotic surgeons from around the United States to include military surgeons. Identify the fundamental knowledge and skills that should be a foundation for every robotic surgeon.

Milestone: FRS consensus conference reports. Award + 180 days and 365 days

Period 2

Telesurgery: Automatic Surgery. Apply movements recorded in a robotic simulator to actual execution with the da Vinci robot on solid models. Explore ability to automatically execute surgery from a simulator recording.

Milestone: Automatic surgery experiment results. Award + 730 days

Simulation: Surgical Rehearsal. Experiment with the effectiveness of simulated surgical rehearsal on improving the outcomes of robotic surgery.

Milestone: Surgical rehearsal experiment results. Award + 540 days

FRS Curriculum Validation and Transition. Develop specific training tasks and passing criteria for the FRS curriculum. Process the curriculum through the certifying bodies.
Milestone: Telesurgery medical procedure results. Award + 730 days

Phase 2: Awarded Extension SOW

Telesurgery: Metropolitan Latency. Perform robotic surgical experiments between multiple campuses within a metropolitan area, between campuses across a state area, and across nationwide campuses.

Period 1 Milestone: Telesurgery state-wide latency data and report. Award + 360 days.
Period 2 Milestone: Telesurgery nationwide latency data and report. Award + 700 days.

Surgical Rehearsal. Develop virtual reality environment for training operating room staff in robotic surgery. Develop design for simulators in hard-tissue robotic surgery (spinal and orthopedic).

Period 1 Milestone: Spinal simulator design document. Award + 300 days.
Period 2 Milestone: OR team training virtual world environment. Award + 360 days.
Period 2 Milestone: Orthopedic surgery rehearsal validation report. Award + 720 days.

Evaluating Simulator Metrics. Compare the metrics assigned by expert surgeons to those assigned by the simulator software.

Milestone 1: Simulator Metric Evaluation Document. February 28, 2107.

Project Management

Progress Summary.

All of the research studies and device prototypes called for in the SOW have been completed and delivered to the government. Some projects resulted in prototypes which have been matured into commercial products through additional investment by Florida Hospital, or they are digital online products which have been released free of charge on the internet and maintenance expenses are being paid by Florida Hospital to continue to make them available to the public.

No negative or adverse event occurred during the course of the work.

Schedule.

During the terms of both the original and the extension, we requested a no-cost extension of the period to complete the scientific work. In both cases, we needed additional time due to either (a) longer than expected staffing times at the beginning, or (b) need to coordinate with other organizations' schedules. In both cases, this meant we spent money slower than planned. So we did not need additional funds, simply additional time in which to use the funds already awarded. The government kindly granted both requests for NCE.

The abbreviated table below shows the completion times for each of the projects performed under the study.

Category	**Project**	**Completion**
Robotic Curriculum		
	Online Curriculum	Mar 2014
	Psychomotor Dome	June 2014
	Validation Pilot	Aug 2014
Simulator Evaluations		
	Surgical Rehearsal	April 2015
	Maintenance of Surgical Skills	Nov 2015
	Simulator Performance	Nov 2015
	Evaluation of Simulator Metrics	Feb 2017
Robotic Simulator Design		
	OR Virtual World	June 2016
	Spinal Robotics	Feb 2017
	Orthopedic Robotics	Feb 2017
Telesurgery		
	Communication Latency	Feb 2015

Budget.

All projects were completed within the allotted budget of the agreement.

Scientific Progress

Robotic Curriculum

We developed the *Fundamentals of Robotic Surgery (FRS),* a surgeon-defined curriculum for teaching the basics of robotic surgical skills. This curriculum development project led to three products:

Online Curriculum. The knowledge-based curriculum of FRS has been completed and posted as an interactive online curriculum. This curriculum is open for use by anyone in the world and can be accessed at: www.FRSurgery.org.

Psychomotor Skills Dome Prototype. The psychomotor skills required by FRS are demonstrated using a physical device which was prototyped under this agreement. That prototype was then carried through full product development by additional investment by Florida Hospital. The resulting product is available as a product and can be found at www.FRSdome.com.

Validation Pilot Study. We subjected the curriculum and device to a multi-site validation trial at 14 different locations around the world. Funds from this agreement were used to carry out a pilot of this very complex trial. The multi-site trial itself was then carried out though education grants from other sources. But that trial could not have been carried out without the pilot study which was conducted by Florida Hospital under this funding. The results of the validation trail will appear as a journal article in 2017.

The published papers describing this work are included as appendices of this report. Additional details on the progress of the work can be found in the annual reports submitted throughout the term of this cooperative agreement.

Simulator Evaluations

Surgical Rehearsal. We conducted a study which compared the effectiveness of the dV-Trainer simulator as a substitute for traditional lecture and video methods of teaching a student to close an incision using a running suture with the da Vinci robot. The results showed that simulator-based instruction was equivalent to lecture and video. We were not able to demonstrate superiority of simulator-based instruction, which we believe was because the procedure performed was too simple to create differing levels of competence. We selected the incision closure for the experiment because it was the only tissue-based procedure represented in the simulator at the time.

Maintaining Surgical Skills. We compared the ability of multiple simulators of the da Vinci robot to improve the performance in specific skills, and the usefulness of these devices to surgical instructors in teaching this information and skills.

Simulator Performance. We explored the ability of different populations to use a simulator effectively to perform robotic skills. The populations compared were experienced surgeons, expert video gamers, medical students, and lay persons. We found that, contrary to generally accepted assumptions, the expert video gamers did not demonstrate a skill level superior to lay

persons and medical students, or even remotely similar to expert surgeons. Therefore, it appears to be false that mastery of video games confers skills which are also applicable to robotic surgery.

Evaluation of Simulator Metrics. Robotic surgery performance is generally evaluated using one of two methods: (1) a human instructor observing a performance and scoring the performance using the GEARS criteria, and (2) a simulator measuring instrument with camera movement, with objective scores applied to the movements. Our experiment applied both methods to a set of subjects and then compared the correlation between the two methods. The results identified subjective metrics in GEARS which are well aligned with the objective metrics in the simulators. It also identified metrics in both areas which could not be correlated with a metric from the other. Therefore, equivalence of the two methods of measuring performance does exist for specific metrics.

Robotic Simulator Design

da Vinci Operating Room Virtual World. We created an online virtual world in which a robotic surgeon can practice the communication and teamwork principles of TeamSTEPPS in a simulated OR environment. In this environment, the rest of the OR team is played by intelligent, computer-driven, avatars. This virtual world has been posted to the internet for anyone in the world to use and Florida Hospital continues to pay the costs required to host and deliver it to interested users. The world can access at: www.TrainRobotic.com.

Spinal Robotic Simulator. A number of additional robotic assistance surgical devices have emerged other than the da Vinci. However, no simulators exist for those devices. We developed a simulator design document for a simulator to be used with the Mazor Renaissance spinal surgery robot. That design document was submitted with the prior quarterly report.

Orthopedic Robotic Simulator. We have developed a simulator design document for the Mako Rio hip & knee orthopedic surgery robot. That design document was submitted with the prior quarterly report.

Telesurgery

We conducted two forms of experiments in telesurgery. The first used the dV-Trainer simulator to explore the levels of communication latency that could be tolerated by a surgeon, specifically because it created a delay between the surgeon's actions and their perception of those actions. In this experiment we determined that: (a) latency between 0 and 250 milliseconds was not perceptible by the surgeon and therefore probably completely safe; (b) latency between 250 and 500 milliseconds was perceptible to all surgeons, but most were able to adjust their actions to compensate and safely complete the procedure; and (c) latency between 500 and 1,000 milliseconds was so extreme that most surgeons could not compensate, could not complete the exercise, and is unsafe for patients. The second form of the experiment was to measure the data delivery speeds on the existing network infrastructure within and between modern hospitals in the United States. Our measurements found that communication latency between multiple campuses of the same hospital system (Florida Hospital) within a large metropolitan area (physical distance < 26 miles) was very steady at approximately 5 milliseconds. Communication latency between hospitals across a statewide area (Florida) delivered latency levels between 10

and 150 milliseconds. Communication latency across states (Florida to Texas and Colorado) also delivered latency levels of approximately 150 milliseconds. Finally, previous research data collected from University of Washington indicated that latency Florida-to-Washington could be as fast as 75 milliseconds (though the actual physical experiment could not be conducted).

Key Research Accomplishments

- *Fundamentals of Robotic Surgery*. We have created a surgeon-defined curriculum for teaching robotic surgery and shared it with the world.
- *Telesurgery: Communications Latency.* Major hospital systems have sufficient telecommunication bandwidth to perform robotic telesurgery right now.
- *Simulator Performance.* The metrics used within robotic surgery simulators is equivalent to the scoring performed by human instructors.
- *Video Game Skills.* Video gamers do not develop skills which are directly applicable to robotic surgery exercises.
- *Simulator Design.* Simulation-based training for different forms of robotic procedures appears to be feasible beyond the simulators of the da Vinci robot which have previously been created. These could be applied to systems like the Mazor Renaissance spinal robotic system and the Mako Rio orthopedic robotic system.

Reportable Outcomes

Refereed Publications:

1. Julian, Tanaka, Mattingly, Perez, Truong, Simpson, Smith. "Comparative Analysis of Four Simulators of the da Vinci Surgical Robot", *American Journal of Surgery*, (Under review).

2. Julian, Tanaka, Mattingly, Smith. (Dec 2016). "Surgical Simulator Showdown." *2016 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*.

3. Smith, Tanaka, McIllwain, Willson. (Dec 2015). "Developing Game-based Leadership Training for Robotic Surgeons." *2015 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*.

4. Tanaka, Graddy, Smith, Perez. (Dec 2015). "Gamers Today, Surgeons Tomorrow?" *2015 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*.

5. Tanaka, Graddy, Simpson, Perez, Truong, & Smith. (2015) "Robotic Surgery Simulation Validity and Usability Comparative Analysis". *Journal of Surgical Endoscopy*.

6. Tanaka, Perez, Truong, & Smith. "From Design to Conception: An Assessment Device for Robotic Surgeons", *2014 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*. December 2014. *Best Paper Nominee*

7. Tanaka, Graddy, & Smith. "Comparison of the Usability of Robotic Surgery Simulators", *2014 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*. December 2014. *Honorable Mention for Best Paper*

8. Smith, Truong, & Perez. (2014) Comparative analysis of the functionality of simulators of the da Vinci surgical robot. *J Surg Endosc*, 1-12.

9. Perez, Xu, Chauhan, Tanaka, Simpson, Abdul-Muhsin, & Smith. "Impact of delay on telesurgical performance: Study on the dV-Trainer robotic simulator". Submission to *Journal of Urology*, 2014.

10. Smith, Patel, & Satava. "Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development", *The International Journal of Medical Robotics and Computer Assisted Surgery*, October 2013. DOI: 10.1002/rcs.1559

11. Smith, "From FLS to FRS: The Fundamentals of Robotic Surgery are on their Way", World Robotic Gynecologic Congress, Chicago, IL. 2013

12. Smith & Truong. "Robotic Surgical Education with Virtual Simulators", *2013 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*. December 2013.

13. Smith & Chauhan. “Using Simulators to Measure Communication Latency Effects in Robotic Telesurgery”, *2012 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*. December 2012 *Best Paper Nominee*

Book Chapters:

1. George, Smith, Levy, Brand, “Robotic Surgery Simulation”, *Simulation for Surgery and Surgical Subspecialities*, Springer, 2017 (in press)
2. Smith, Tanaka, Julian, Mattingly, “Enseñanza Semipresencial para la Educación en Cirugía Robótica”, *Simuladores Quirurgicos y Cirugia Robotica*, Universidad of Panamericana Press, 2017.
3. Smith, Tanaka, Perez, Truong, Simpson, “Fundamentos de Cirugía Robótica: Proceso de diseño y método de evaluación para habilidades psicomotoras”, *Simuladores Quirurgicos y Cirugia Robotica*, Universidad of Panamericana Press, 2017.

Conference Presentations:

1. Julian, Tanaka, Mattingly, Smith, “Surgical simulator showdown”, *Proceedings of the 2016 Interservice/Industry Training Education and Simulation (I/ITSEC) Conference*.
2. Tanaka, Smith, Hughes, “Video Game Experience and Basic Robotic Skills”, IEEE International Conference on Serious Games and Applications for Health, May 2016.

3. Mattingly, Tanaka, Julian, Skinner, Smith. (Nov 2016). “Simulator-based Multi-modal Task Decomposition of Robotic Surgical Technique for Vaginal Cuff Closure.” Annual Meeting of the American Association of Gynecologic Laparoscopists.

4. Smith, “Robots in Surgery and Simulation in Training.” IEEE International Systems Conference, April 2016.

5. Julian, Tanaka, Smith, “A Side-by-Side Comparison of Virtual Reality Robotic Surgical Simulators.” Florida Hospital Research Conference, April 2016.

6. Tanaka, Smith, Hughes, “Video Game Experience and Basic Robotic Skills.” IEEE International Conference on Serious Games and Applications for Health, February 2016.

7. Smith R. “The Validation of Surgical Simulators for RASD”. FDA Workshop on Robotically Assisted Surgical Devices, Washington DC, August 2015.

8. Tanaka, Graddy, Perez, Simpson, Truong, Smith. (Nov 2015). “Video Game Impact on Basic Robotic Surgical Skills.” Annual Meeting of the American Association of Gynecologic Laparoscopists.

9. Perez, Tanaka, Simpson, Truong, Smith, Satava. (Nov 2015). “From concept to surgical relevance: Engineering the training device for the Fundamentals of Robotic Surgery.” Annual Meeting of the American Association of Gynecologic Laparoscopists.

10. Perez M, Tanaka A, Bresler L, Hubert J, Smith R. Invitée à présenter à l'Académie Nationale de Chirurgie. Séance du mercredi 20 mai 2015 (SÉANCE DÉLOCALISÉE À NANCY / Apprentissage des différentes techniques chirurgicales par simulation) Fundamentals of Robotic Surgery: Future certification en chirurgie robotique aux USA: Un avenir en France?
11. Tanaka, Perez, Graddy, & Smith. "Video Game Experience and Basic Robotic Skills", Florida Hospital Internal Research Forum, Orlando, FL, April 2015.
12. Truong, Tanaka, Simpson, Perez, & Smith. "Robotic surgical simulation versus traditional didactics for surgical training: a randomized controlled trial", Society for Gynecologic Surgeons Annual Scientific Meeting, Orlando, FL, March 2015.
13. Smith. "Update on Robotic Surgical Simulation", *2015 Society of Robotic Surgeons (SRS),* Orlando, FL, February 2015.
14. Smith. "Fundamentals of Robotic Surgery", *2015 Society of Robotic Surgeons (SRS),* Orlando, FL, February 2015.
15. Truong, Tanaka, Simpson, Advincula, & Smith. "A Prospective Randomized Controlled Comparative Study on Surgical Training Methods and Impact on Surgical Performance: Virtual Reality Robotic Simulation vs. Didactic Lectures", AAGL Global Congress on Minimally Invasive Gynecology, November 2014
16. Tanaka, Truong, & Smith. "Robotic Surgical Simulators: An Assessment of Usability and Preferences", AAGL Global Congress on Minimally Invasive Gynecology, November 2014
17. Simpson, Perez, Tanaka, Truong & Smith. "Validating the Efficacy of GEARS through the Assessment of 100 Videos", Society of Laparoendoscopic Surgeons Annual Meeting & Endo Expo, September 2014.
18. Truong, Tanaka, Simpson, Perez, Smith & Advincula. "Randomized Controlled Study Comparing Robotic Simulation Versus Didactic Teaching for Robotic Surgical Training: Opinions and Perspectives", Society of Laparoendoscopic Surgeons Annual Meeting & Endo Expo, September 2014. *Honorable Mention for the Paul Alan Wetter Award for Best MultiSpecialty Scientific Paper*
19. Smith & Tanaka. "Gamers in Surgical Simulation: A Comparison of Gamers, Surgeons, and Clinical Staff", Defense GameTech Users Conference, Orlando, FL, September 2014.
20. Lendvay, Simpson, Truong, & Smith. "Differentiating Surgical Skill through the Wisdom of Crowds", European Endoscopic Urology Society, April 2014.
21. Patel, Patel & Smith, "Feasibility of Robotic Telesurgery across a Multi-Campus Metropolitan Hospital System", Third Biennial Miami Robotics Symposium, April 2014.
22. Smith, "Robotic & Telesurgery Research", Stetson University Senior Tech Expo, March, 2014.

23. Satava & Smith, “Fundamentals of Robotic Surgery: Development and Validation of an Online Curriculum and New Psychomotor Testing Device”, NextMed/MMVR Conference, February, 2014.

24. Satava & Smith, “Fundamentals of Robotic Surgery: Development and Validation of an Online Curriculum and New Psychomotor Testing Device”, CAMLS-Halldale Summit on New Technology in Medicine, February, 2014.

25. Tanaka, Truong, Simpson, Perez, & Smith, “A Comparison of the Effectiveness and Usability of Robotic Simulators”, Florida Hospital Internal Research Forum, January 2014.

26. Truong: "The Fundamentals of Robotic Surgery Psychomotor Skills Prototype Development Video": Harrith M Hasson Award for Best Presentation Promoting Education and Training, 2013 SLS Annual Meeting in Reston, Virginia. Smith, “Robotic Surgery Education, Simulation & Telesurgery, Society for Laparoscopic Surgeons, Fellowship Summit, December 2013.

27. Smith, “Virtual Reality Simulation: The Future”, Society for Robotic Surgery, Annual Meeting, November, 2013.

28. Smith, “Robots in the Hands of Your Surgeon”, Chinese American Scholars and Professionals Association of Florida, Miami Annual Meeting, August 2013.

29. Smith, “Robotic Surgery and Surgical Simulation”, presentation to *International Council on Systems Engineering – Orlando Chapter*. February 2012.

30. Smith, “Beyond Education and Training: Challenges of Running Medical Simulators in New Paradigms”. *2012 International Meeting on Simulation in Healthcare*. January 2012.

31. Smith, "Simulation in Surgical Education", American College of Healthcare Executives, December 2011.

32. Smith, "Medical Simulation Special Event: Robotic and Telesurgery Research Using Simulation", I/ITSEC, December 2011.

33. Smith, "Robotic and Telesurgery Research", National Center for Simulation, October 2011.

34. Smith, "Robotic Surgery and Surgical Simulation", Guest Lecture, Old Dominion University, April 2011.

35. Smith, "Surgical Simulation Research Initiatives", I/ITSEC UCF Workshop, December 2010.

Poster Presentations:

1. Mattingly, Tanaka, Julian, Smith. (2016) Virtual Reality Robotic Simulation Performance Assessment: Simulator Metrics vs. Global Evaluative Assessment of Robotic Skills (GEARS). Annual Meeting of the American Association of Gynecologic Laparoscopists.

2. Julian, D., Tanaka, A., Mattingly, P., & Smith, R. (2016). "A Side-by-Side Comparison of Virtual Reality Robotic Surgical Simulators." University of Central Florida Graduate Research Forum 2016. Orlando, Fl.

3. Tanaka, Graddy, Perez, Simpson, Truong, Smith. (Jan 2016). "Video Game Impact on Basic Robotic Surgical Skills." International Meeting on Simulation in Healthcare, San Diego, CA.

4. Smith, Simpson. "Return on Investment Model for Robotic Simulators", Poster Presentation at *2015 Society of Robotic Surgeons (SRS),* Orlando, FL, February 2015.

5. Tanaka, Graddy, Abdul-Muhsin, Simpson, Truong, & Smith. "A Comparison of Validity and Usability of Robotic Simulation", Poster Presentation at *2015 Society of Robotic Surgeons (SRS),* Orlando, FL, February 2015.

6. Tanaka, Perez, & Smith. "Fundamentals of Robotic Surgery Psychomotor Skills: Metrics Development and Evaluation", Poster Presentation at *2015 International Meeting on Simulation in Healthcare (IMSH) Conference*, New Orleans, LA, January 2015.

7. Lendvay TS, White LW, Holst D, Kowalewski T, Harper JD, Sorenson M, Brand TC, Truong M, Simpson K, Smith R. Quantifying Surgical Skill Using the Wisdom of Crowds. *American College of Surgeons Clinical Congress*, San Francisco, CA, October 26-30th, 2014. [Poster #PP2014-51161].

8. Lendvay T, Holst D, White L, Kowalewski T, Brand T, Sorenson M, Harper J, Truong M, Simpson K, Smith R. Differentiating Surgical Skill Through the Wisdom of Crowds. *American Urological Association Annual Meeting*, Engineers in Urology Session, Orlando, FL, May 16-21, 2014 [Moderated Poster #82].

Conclusion

Each of the research areas funded by this grant has made significant scientific contributions. The knowledge gained from this work is being shared through reports to the government, journal publications, and multiple presentations at both clinical and simulation conferences. The digital products of the research have also been made freely available to use across the world. Surgeons and surgical instructors have free access to the educational materials developed for the Fundamentals of Robotic Surgery (FRS) and to the online virtual world created. FRS is on a track to become an international standard for education and performance measurement for all practitioners of robotic surgery. The work was conducted at a time when there was one primary robotic system (da Vinci), however, we are in discussions with multiple companies who are releasing new robots who want to adapt and apply this curriculum to their devices as well.

It has been a privilege to conduct this research and make these contributions to the scientific literature and to those teaching and advancing work in robotic assisted surgery.

References

(No external sources are referenced in this report. All information is derived from the list of our own publications in the Reportable Outcomes section. External references are cited in those publications.)

Bibliography

All publications and presentations resulting from this work are included in the Reportable Outcomes section, per the instructions in "USAMRAA Technical Reporting Requirements (Dec 2008)" which is included in the award document.

Personnel

The following staff of Florida Hospital received salary support from this cooperative agreement.

Scientific Staff:

- Roger Smith, PhD
- Alyssa Tanaka, PhD
- Danielle Julian, MS
- Ariel Dubin, MD
- Patricia Mattingly, MD
- Manuela Perez Santalices, MD
- Khara Simpson, MD
- Mireille Truong, MD
- Haidar Abdul-Mushin, MD
- Sanket Chauhan, MD

Miscellaneous Support Staff:

- Anthony Basica
- Manuel Caday
- Edith Castro
- Holly Daniel
- Courtney Graddy, MS
- Gareth Hearn, MS
- Tamara Lecrone
- Norma Neives
- Monique Palmore
- Steven Thekan
- John Venturelli

Appendices

Copies of manuscripts, abstracts, and presentations of work resulting from this grant are included as appendices to this report.

Current status of robotic simulators in acquisition of robotic surgical skills

Anup Kumar[a], Roger Smith[b], and Vipul R. Patel[a]

Purpose of review
This article provides an overview of the current status of simulator systems in robotic surgery training curriculum, focusing on available simulators for training, their comparison, new technologies introduced in simulation focusing on concepts of training along with existing challenges and future perspectives of simulator training in robotic surgery.

Recent findings
The different virtual reality simulators available in the market like dVSS, dVT, RoSS, ProMIS and SEP have shown face, content and construct validity in robotic skills training for novices outside the operating room. Recently, augmented reality simulators like HoST, Maestro AR and RobotiX Mentor have been introduced in robotic training providing a more realistic operating environment, emphasizing more on procedure-specific robotic training . Further, the Xperience Team Trainer, which provides training to console surgeon and bed-side assistant simultaneously, has been recently introduced to emphasize the importance of teamwork and proper coordination.

Summary
Simulator training holds an important place in current robotic training curriculum of future robotic surgeons. There is a need for more procedure-specific augmented reality simulator training, utilizing advancements in computing and graphical capabilities for new innovations in simulator technology. Further studies are required to establish its cost–benefit ratio along with concurrent and predictive validity.

Keywords
robotics surgery, simulation, surgical training, virtual reality

INTRODUCTION

The use of the robotic platform in urology has expanded exponentially over the last decade and has established itself in most advanced centres across the world, particularly in the USA [1–3]. In 2013, approximately 80% of radical prostatectomies were performed using robotic platform in the USA [1]. This tremendous growth in robotic technology has highlighted the increasing demand for surgeons trained in robotic skills. Although most urology residency programs are presently incorporating robotic surgery as a part of their curriculum, adequate training of these future robotic surgeons is facing many challenges [4–6]. First, there has been a decrease in actual training hours along with risk of litigation, increased emphasis on patient safety and improved surgical outcomes. Second, the traditional Halstedian method of training of 'see one, do one and teach one' does not apply to robotic technology. The robot-assisted radical prostatectomy is a complex procedure requiring complete knowledge of pelvic anatomy and an understanding of magnification, depth perception, three-dimensional spatial orientation and coordinated hand–eye movements. Third, in robotics, the mentor is not working close to the trainee with one person at the console and one other person required for bedside assistance, thus raising concerns in the mentor's mind about the patient's safety [7–9]. The training can be divided as preclinical and clinical [4–6]. The preclinical training includes use of simulators, defined as tools

[a]Department of Urology, Global Robotics Institute, Florida Hospital-Celebration Health and [b]Chief Technology Officer, Florida Hospital-Nicholson Center, University of Central Florida College of Medicine, Orlando, Florida, USA

Correspondence to Dr Anup Kumar, Department of Urology, Global Robotics Institute, Florida Hospital-Celebration Health and University of Central Florida College of Medicine, Orlando, FL 32827, USA. Tel: +1 407 701 0365; e-mail: anup_14k@yahoo.com

Curr Opin Urol 2015, 25:168–174

DOI:10.1097/MOU.0000000000000137

KEY POINTS

- The simulator training can form an integral part of credentialing and training robotic surgery of future robotic surgeons.
- It has the potential to decrease the learning curve for the acquisition of robotic skills.
- It can supplement the hands-on training clinical phase and can act as a bridge between preclinical training and actual hands-on clinical training without jeopardizing the safety of patients.
- There is a need for more procedure-specific augmented reality simulator training in a cost-effective manner, with more emphasis on both technical skills and teamwork training.

enabling the operator to reproduce or represent under test conditions a phenomenon likely to occur in actual performance. Clinical training includes observation, bed-side assistance and hands-on-training under mentorship (including Tele-mentoring) and proctoring [4,7,10].

Simulators can be classified as low fidelity, high fidelity, virtual reality and augmented reality [4–7,11,12▪▪]. Low fidelity simulators, like Dry lab laparoscopic box trainer, are portable, less expensive and have been proven to improve surgical skills over time. But, they have disadvantages of lack of duplication of a real surgical environment, lack of feedback and inability to teach an entire procedure. High fidelity simulators include animal models, cadavers and commercially available models. They have advantages of providing a more realistic environment for training, but also have disadvantages such as lack of easy availability, cost, ethical issues, veterinary assistance, anatomical variance from human organs (with animal models) and lack of bleeding and actual tissue compliance (for cadavers). The Virtual Reality simulator utilizes a computer-derived realistic virtual operative field with tactile feedback on laparoscopic instruments. The Augmented Reality simulator provides a more realistic procedure-specific operating environment, where events on the field are enhanced and supplemented [12▪▪,13,14].

Simulators enable residents and novice robotic surgeons to practice their skills in a nonclinical environment, any number of times, without risking the actual patients. Moreover, they provide trainees a platform to assess their performance and keep track of progress over time. Additionally, they provide an opportunity to a surgeon to refamiliarize himself with the surgical console immediately before a case as a 'warm-up' before surgery [4–10].

The simulator training can be further classified into two types – skills training and procedure-based training [4–6]. Most of the virtual reality simulators provide skills training including cutting, depth perception, hand–eye coordination, suturing and retraction. Recently, procedure-based training simulators have been reported, which can act as a bridge between formal and informal training [13,14,15▪▪].

In this systematic review, we have reviewed all publications in PubMed in the last 12 months using keywords: simulation, robotic training, virtual reality, augmented reality. We will discuss the current status of all existing simulators in robotic training including their advantages, disadvantages, all recently published modifications in simulators technology, assessing their place in current robotic training curriculum, along with the recent developments in simulator technology and future challenges in the simulator training for acquisition of robotic skills.

VALIDATION OF SIMULATORS

Although simulators have shown their utility over other educational tools like didactic teaching and dry lab training, they need to be validated before their effective integration into teaching and training curriculum [4–6]. Validation can be subjective and objective. The subjective validation includes face and content validity. Face validity is defined as the informal assessment of realism and feel by no experts. Content validity is defined as the formal assessment of appropriateness as a teaching tool by experts. The objective validation, which is a much more daunting task, includes construct, concurrent and predictive validity. Construct validity is defined as the ability of a simulator to discriminate experts from novices. The term 'novice' includes subjects with no experience at all in performing the procedure under study. The term 'expert' includes subjects with adequate experience in performing the procedure under study. Concurrent validity is defined as the ability to compare performance on a simulator with gold standard tests known to measure the same domain, such as a tissue or animal lab. Predictive validity is defined as the ability to predict future performance based on performance on the simulator [4–10].

VIRTUAL REALITY SIMULATORS

We found five different types of virtual reality simulators published so far in the literature.

SimSurgery Educational Platform Robot

The SimSurgery Educational Platform (SEP) Robot (SimSurgery, Oslo, Norway) is a modification of the SEP Basic laparoscopic virtual reality simulator. It replaces the simulated laparoscopic instruments with the wristed instruments found in the da Vinci robot, providing seven degrees of freedom. It does not provide three-dimensional images, fourth arm integration or performance feedback. It also does not include the following tasks: camera and clutching; needle control and driving; energy and dissection [9,16]. The experience with this simulator is not as robust as with other simulators, though it is an extremely cost-effective alternative. However, the face, content and construct validity have been proven in literature [9,16].

Robotic Surgical Simulator

The Robotic Surgical Simulator (RoSS) is another type of virtual reality simulator offering 16 modules with progressive difficulty from pinching, camera and clutch operation to tissue cutting and cautery. It is a stand-alone system mimicking da Vinci Surgical System. It helps in developing motor and cognitive skills for performing robotic surgery by providing in-vivo virtual operative steps with three levels of complexity in the form of modules for orientation, motor skills, basic surgical skills and intermediate surgical skills [17]. The face and content validity have been published for this simulator, but there is currently no literature on construct validity [4,9]. The educational impact of this simulator has been published as those trained on RoSS took less time to complete robotic dry tasks [18]

ProMIS

The ProMIS hybrid simulator (Canadian Aviation Electronics Healthcare, Canada) has a computer and a laparoscopic interface made with a plastic mannequin with a black Neoprene cover. There are three camera tracking systems to detect any instrument inside the simulator from three angles, thus recording the three-dimensional position of tips of instruments 30 times/second. It can be used for various tasks like intracorporeal suturing, precision cutting, cannulation and peg transfer, analyzing three objective parameters of time, path and smoothness [19]. The face, content and construct validity have been reported in published literature [9,19].

Mimic dV-Trainer

dV-Trainer (dVT) is a table top-sized compact system with dual-platform capability simulating both da Vinci S, Si and Xi robots. It utilizes precise modelling of robot kinematics, foot pedals and master grips. This provides trainees with a realistic representation of the da Vinci system. This provides both basic (Endowrist manipulation, camera, clutching, and troubleshooting) and advanced skills training (needle control and driving, suture and knot tying, energy and dissection) [4,7]. The face, content, construct validity and educational impact have been proven in recent published series [6,18,20–22]. Schreuder *et al.* evaluated 42 participants in three groups according to their robotic experience. Experts performed better in terms of 'time to complete' and 'economy of motion' in comparison to novices [20].

da Vinci Skills Simulator

This simulator, produced by Intuitive Surgical, can be integrated with existing da Vinci Xi or Si surgeon consoles, thus providing a practice platform to be used inside or outside the operating room, with no requirement of additional system components. This was developed in collaboration with Mimic Technologies and Simbionix and provides training modules from basic to advanced skills including Endowrist manipulation, camera and clutching, fourth arm integration, needle control and driving, energy and dissection [4,23]. The face, content and construct validity have been proven in the recent series [11,18,24–28]. Tergas *et al.* showed that training on da Vinci Skills Simulator (dVSS) resulted in significant improvement in 'time to completion' and 'economy of motion' for novices [24]. They found that autonomy of use, computerized performance feedback and ease of setup were unique advantages to dVSS, thus providing more efficient and sophisticated training in comparison to conventional dry laboratory training.

AUGMENTED REALITY SIMULATORS

These simulators provide a more realistic operating field to trainees, utilizing enhanced and supplemented events [29].

Hands-on-Surgical Training

This simulator is a mode embedded within the RoSS simulator and provides training in actual surgical cases such as radical prostatectomy, radical cystectomy, radical hysterectomy and extended lymph node dissection. It includes integrated user interaction, narrative instructions and guided movements. Hands-on-Surgical Training (HoST) was created by augmenting a real surgical procedure

within a virtual reality framework utilizing audiovisual explanations and anatomically relevant illustrations of the critical steps of the procedure. The RoSS manipulators navigate the trainee through haptic-enabled cues during the procedure [13]. Chowriappa *et al.* [12▪▪] evaluated the role of augmented reality-based skills training for robot-assisted urethrovesical anastomosis in a randomized controlled trial, using HoST a technology group and a control group. They found that for 70% of participants, HoST the training experience was similar to a real surgical procedure and 75% of trainees responded that this training could improve confidence in performing a real procedure. They concluded that training with HoST in urethrovesical anastomosis improves technical skills acquisition with minimal cognitive demand.

Maestro AR

This was introduced by Mimic Technology, providing virtual instruments for interaction with anatomy in a 3D video environment. This has been designed for training novices in decision-making skills and procedure-specific skills, within the dVT simulator. The participants use virtual robotic instruments in anatomical regions collected from 3D surgical video. This simulator plans to provide training in four modules: partial nephrectomy (released May 2014), hysterectomy, prostatectomy and general surgery (to be released) by helping to identify anatomy, anticipate tissue retractions and predict regions for dissection [14]. There are no studies documenting face, content, construct, concurrent and predictive validity of this simulator, owing to its recent introduction.

RECENT DEVELOPMENTS IN CONCEPTS

Recently, more simulation models have been launched emphasizing the concept of teamwork and procedure-specific training in robotics.

Xperience Team Trainer

This simulator, available as an optional hardware complement for the dV–Trainer simulator, has been introduced to emphasize the importance of teamwork and proper coordination between console surgeon and assistant during robotic surgery. This simulator provides training simultaneously to both surgeon and bedside assistant. Thus, the bedside assistant performs basic skills exercises, promoting his psychomotor skills and rehearsal of interaction with console surgeon. It also exposes them to real-life situations in the operating room, promoting patient safety. Moreover, this team training helps in development of communication protocol in the real operating room using a well tolerated simulation environment. Moreover, it also provides proficiency-based scoring for the team and each individual [30]. However, studies regarding its face, content, construct, concurrent and predictive validity are still pending because of its recent introduction.

Tube 3 module with dV-Trainer

This simulator training emphasizes procedure-specific training, utilizing the Tube 3 module in the dVT. It helps in increasing vesicourethral anastomosis (VUA) performance, one of the most complex steps in robot-assisted radical prostatectomy. Kang *et al.* [15▪▪] recently published their experience with this module. They found that experts performed better in task time, total score, total economy of motion and number of instrument collisions in comparison with novices. Moreover, 80% of experts found this module a useful training tool to perform VUA. Thus, they reported face, content and construct validity of the Tube 3 module for practicing VUA.

RobotiX Mentor

This simulator has been introduced recently providing a realistic representation of the work space, master controllers, pedals and surgeon console of da Vinci Surgical System. It provides a 3D high-definition stereoscopic view for basic skills (robotic suturing, stapler, Fundamentals of Robotic Surgery modules) and multidisciplinary complete virtual reality procedures (vaginal cuff closure, hysterectomy modules), augmented with step-by-step video guidance and realistic representation of emergency situations and complications. The trainees are provided with performance reports with learning curve graphs utilizing simulator curricula management system [31]. However, face, content, construct, concurrent, and predictive validity of this simulator have not been proved in literature because of its recent introduction.

Table 1 shows comparison between the available simulators.

CURRENT CHALLENGES AND FUTURE PERSPECTIVES

The definitions of face, content, construct, concurrent and predictive validity need to be standardized for all simulators and future studies. Very few randomized controlled trials (RCTs) have been

Table 1. Comparison of different available simulators

	Face validity	Content validity	Construct validity	Concurrent validity	Predictive validity	Learning impact	Cross-modality correlation
SEP	Yes	Yes	Yes	No	No	No	No
RoSS	Yes	Yes	No	No	No	Yes	No
ProMIS	Yes	Yes	Yes	No	No	Yes	No
dVT	Yes	Yes	Yes	No	No	Yes	Yes
dVSS	Yes	Yes	Yes	No	No	Yes	Yes
HoST	Yes	Yes	Yes	No	No	Yes	No
Maestro AR	No	No	No	No	No	No	No
Tube-3 module	Yes	Yes	Yes	No	No	Yes	No
Xperience team trainer	No	No	No	No	No	No	No

dVSS, da Vinci Skills Simulator; dVT, dV-Trainer; HoST, Hands-on-Surgical Training; RoSS, Robotic Surgical Simulator; SEP, SimSurgery Educational Platform.

reported comparing different robotic simulators [32]. The superiority of one simulator over another has not been established so far because of a lack of these RCTs. There are no studies documenting the actual benefits of simulator training carried over to real-case performance with a surgical robot. The cost of these simulators is a significant matter of concern [4,7–9]. However, with increasing use of robotic technology and increasing competition among training devices, the future cost of these devices should come down to an affordable range. There is a need to provide more procedure-specific

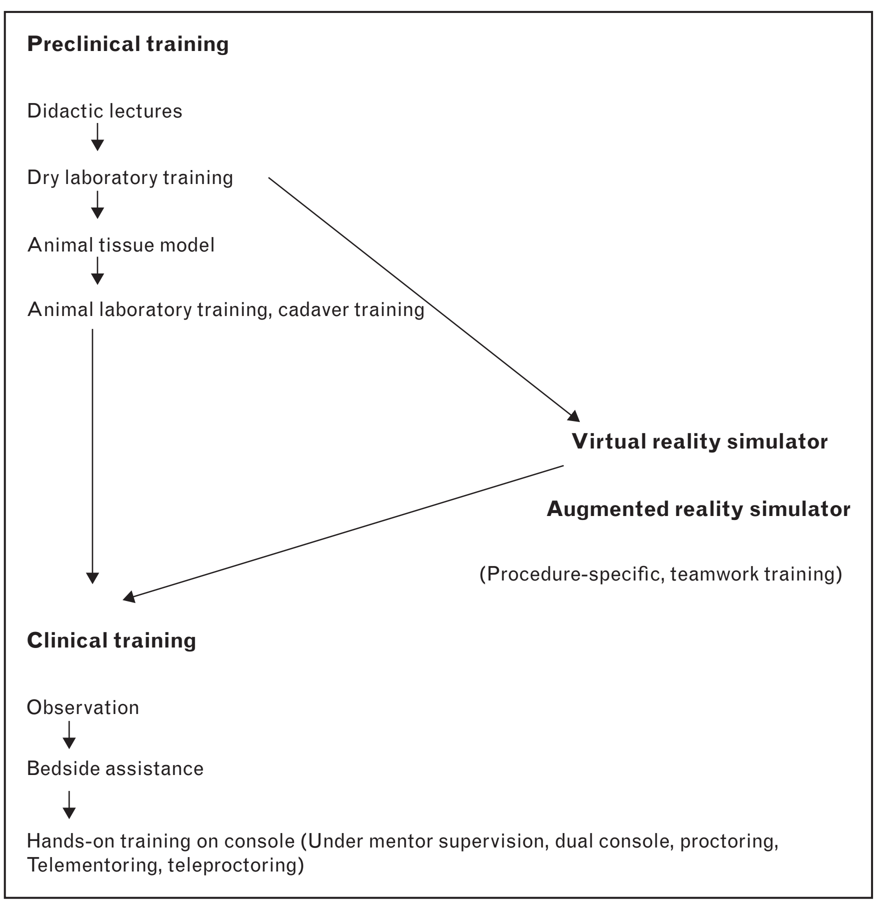


FIGURE 1. Potential role of simulators in robotics training.

training along with skills-based training in a more realistic augmented reality environment like HoST and Maestro [13,14]. Moreover, the concepts of teamworking, decision-making and communication skills should be incorporated more in simulator training by providing team-based robotic simulation environments like Xperience Team Trainer [4,7–9,30]. However, their validations have to be proved in future large prospective RCTs. Finally, there is a need for standardization for training and credentialing in robotic surgery as has been done with Fundamentals of Laparoscopy Surgery for laparoscopy in general surgery [4,7,8]. A similar standard and validated tool including simulator training and other training tools needs to be incorporated in various robotic residency and fellowship teaching curriculum (Fig. 1).

There are a few limitations of this article. First, we may have missed a few articles related to the current topic. Second, we could not discuss certain issues like cost-effectiveness, concurrent and predictive validity (tools to assess the actual benefits of simulator training carried over during real-time robotic surgery), as these issues have not been reported in published series.

CONCLUSION

The simulator training can form an integral part of credentialing and training robotic surgery of future robotic surgeons. It has the potential to decrease the learning curve for the acquisition of robotic skills. It can supplement the hands-on training clinical phase and can act as a bridge between preclinical training (didactic lectures, dry lab training, animal models) and actual hands-on clinical training without jeopardising the safety of patients. There is a need for more procedure-specific augmented reality simulator training in a cost-effective manner, utilizing advancements in computing and graphical capabilities for new innovations in simulator technology, with emphasis on both technical skills training and teamwork training. However, more RCTs involving larger numbers of participants are required to establish its cost-benefit ratio along with concurrent and predictive validity.

Acknowledgements

None.

Financial support and sponsorship

None.

Presentation at a meeting: None.

Conflicts of interest

None.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:
■ of special interest
■■ of outstanding interest

1. Skarecky DW. Robotic-assisted radical prostatectomy after the first decade: surgical evolution or new paradigm. ISRN Urol 2013; 2013:157379.
2. Huang KH, Carter SC, Shih YC, Hu JC. Robotic and standard open radical prostatectomy: oncological and quality-of-life outcomes. J Comp Eff Res 2013; 2 (3):293–299.
3. Bian SX, Hoffman KE. Management of prostate cancer in elderly men. Semin Radiat Oncol 2013; 23 (3):198–205.
4. Sun AJ, Aron M, Hung AJ. Novel training methods for robotic surgery. Indian J Urol 2014; 30 (3):333–338.
5. Khan R, Aydin A, Khan MS, *et al.* Simulation-based training for prostate surgery. BJU Int 2014; doi: 10.1111/bju.12721. [Epub ahead of print]
6. Patel A, Patel M, Lytle N, *et al.* Can we become better robot surgeons through simulator practice? Surg Endosc 2014; 28 (3):847–853.
7. Buchs NC (1), Pugin F, Volonté F, Morel P. Learning tools and simulation in robotic surgery: state of the art. World J Surg 2013; 37 (12):2812–2819.
8. Liss MA (1), McDougall EM. Robotic surgical simulation. Cancer J 2013; 19 (2):124–129.
9. Abboudi H (1), Khan MS, Aboumarzouk O, *et al.* Current status of validation for robotic surgery simulators - a systematic review. BJU Int 2013; 111 (2):194–205.
10. Santomauro M, Reina GA, Stroup SP, L'Esperance JO. Telementoring in robotic surgery. Curr Opin Urol 2013; 23:141–145.
11. Rajanbabu A, Drudi L, Lau S, *et al.* Virtual reality surgical simulators-a prerequisite for robotic surgery. Indian J Surg Oncol 2014; 5 (2):125–127.
12. ■■ Chowriappa A, Raza SJ, Fazili A, *et al.* Augmented –reality based skills training for robot assisted urethrovesical anastomosis: a multiinstituitional randomised controlled trial. BJU Int 2014; doi: 10.1111/bju.12704. [Epub ahead of print]

This is the only randomized controlled trial reporting role of HoST™ in robotic training.

13. Simulated Surgical Systems.HoST What is HoST?2014. http://www.simulatedsurgicals.com/what-is-host.htm. Accessed September 2014.
14. Mimic Simulation Technology:Maestro AR™ http://www.mimicsimulation.com/products/maestro-ar/. Accessed September 2014.
15. ■■ Kang SG, Cho S, Kang SH, *et al.* The Tube 3 module designed for practicing vesicourethral anastomosis in a virtual reality robotic simulator: determination of face, content, and construct validity. Urology 2014; 84 (2):345–350.

This is the only study reporting role of Tube 3 module with dV-Trainer in procedure specific training.

16. Shamim Khan M, Ahmed K, Gavazzi A, *et al.* Development and implementation of centralized simulation training: evaluation of feasibility, acceptability and construct validity. BJU Int 2013; 111 (3):518–523.
17. Simulated Surgical Systems .RoSS™:Hardware and software. http://www.simulatedsurgicals.com/hw-sw.htm. Accessed September 2014.
18. Smith R, Truong M, Perez M. Comparative analysis of the functionality of simulators of the da Vinci surgical robot. Surg Endosc 2014. [Epub ahead of print]
19. Nagendran M, Gurusamy KS, Aggarwal R, *et al.* Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev 2013; 8:CD006575. doi: 10.1002/14651858.CD006575.pub3.
20. Schreuder Henk WR, Persson Jan EU, Wolswijk Richard GH, *et al.* Validation of a novel virtual reality simulator for robotic surgery. ScientificWorldJournal 2014; 2014:507076.
21. Ruparel RK, Taylor AS, Patel J, *et al.* Assessment of virtual reality robotic simulation performance by urology resident trainees. J Surg Educ 2014; 71 (3):302–308.
22. Thiel DD, Patel VR, Larson T, *et al.* Assessment of robotic simulation by trainees in residency programs of the Southeastern Section of the American Urological Association. J Surg Educ 2013; 70 (5):571–577.
23. Intuitive Surgical.da Vinci Skills Simulator. http://www.intuitivesurgical.com/products/skills_simulator/. Accessed September 2014.
24. Tergas AI, Sheth SB, Green IC, *et al.* A pilot study of surgical training using a virtual robotic surgery simulator. JSLS 2013; 17 (2):219–226.
25. Alzahrani T, Haddad R, Alkhayal A, *et al.* Validation of the da Vinci Surgical Skill Simulator across three surgical disciplines:A pilot study. Can Urol Assoc J 2013; 7 (7–8):e520–e529.
26. Brinkman WM, Luursema JM, Kengen B, *et al.* da Vinci skills simulator for assessing learning curve and criterion-based training of robotic basic skills. Urology 2013; 81:562–566.

27. Lyons C, Goldfarb D, Jones SL, *et al.* Which skills really matter? Proving face, content, and construct validity for a commercial robotic simulator. Surg Endosc 2013; 27:2020–2030.
28. Hung AJ, Jayaratna IS, Teruya K, *et al.* Comparative assessment of three standardized robotic surgery training methods. BJU Int 2013; 112:864–871.
29. Stegemann AP, Ahmed K, Syed JR, *et al.* Fundamental skills of robotic surgery: A multiinstitutional randomized controlled trial for validation of a simulation-based curriculum. Urology 2013; 81:767–774.
30. Mimic Simulation Technology:Xperience Team Trainer™ http://www.mimicsimulation.com/products/xperience/. Accessed September 2014.
31. Simbionix™: RobotiX Mentor™. http://simbionix.com/simulators/robotix-mentor/. Accessed September 2014.
32. Korets R, Mues AC, Graversen J, *et al.* Comparison of robotic surgery skill acquistion between DV-Trainer and da-Vinci surgical system: a randomized controlled study. J Urol 2011; 185:e593.

Int J CARS (2016) 11:581–587
DOI 10.1007/s11548-015-1306-y

ORIGINAL ARTICLE

Impact of delay on telesurgical performance: study on the robotic simulator dV-Trainer

Manuela Perez[1,2,3,7] · Song Xu[1,4] · Sanket Chauhan[5,6] · Alyssa Tanaka[3] · Khara Simpson[3] · Haidar Abdul-Muhsin[3] · Roger Smith[3]

Received: 19 May 2015 / Accepted: 21 September 2015 / Published online: 8 October 2015

Abstract
Purpose To determine the impact of communication latency on telesurgical performance using the robotic simulator dV-Trainer®
Methods Surgeons were enrolled during three robotic congresses. They were randomly assigned to a delay group (ranging from 100 to 1000 ms). Each group performed three times a set of four exercises on the simulator: the first attempt without delay (Base) and the last two attempts with delay (Warm-up and Test). The impact of different levels of latency was evaluated.
Results Thirty-seven surgeons were involved. The different latency groups achieved similar baseline performance with a mean task completion time of 207.2 s ($p > 0.05$). In the Test stage, the task duration increased gradually from 156.4 to 310.7 s as latency increased from 100 to 500 ms. In separate groups, the task duration deteriorated from Base for latency stages at delays ≥300 ms, and the errors increased at 500 ms and above ($p < 0.05$). The subjects' performance tended to improve from the Warm-up to the Test period. Few subjects completed the tasks with a delay higher than 700 ms.
Conclusion Gradually increasing latency has a growing impact on performances. Measurable deterioration of performance begins at 300 ms. Delays higher than 700 ms are difficult to manage especially in more complex tasks. Surgeons showed the potential to adapt to delay and may be trained to improve their telesurgical performance at lower-latency levels.

Keywords Telesurgery · Telemedicine/methods · Computer simulation · Robotic simulator · Internet

Grant Information: Funding Source: U.S. Army Telemedicine and Advanced Technology Research Center. Grant #: W81XWH-11-2-0158. The Grant was provided to the Florida Hospital Nicholson Center without personal benefit for the authors.

✉ Manuela Perez
m.perez@chu-nancy.fr

Roger Smith
roger.smith@flhosp.org

1 IADI Laboratory-INSERM-U947, Lorraine University, Allée du Morvan, 54500 Vandoeuvre-les-Nancy, France

2 General and Emergency Surgery Department, University Hospital of Nancy, Avenue du Maréchal de Lattre de Tassigny, 54035 Nancy, France

3 Florida Hospital Nicholson Center, 404 Celebration Place, Celebration, FL 34747, USA

4 Urology Department, University Hospital of Nancy, Allée du Morvan, 54511 Vandoeuvre-les-Nancy, France

5 Center for Evidence Based Simulation, Baylor University Medical Center, Dallas, TX, USA

6 Associate Professor Texas A&M Health Science Center, College Station, TX, USA

7 Hôpital Central Service de Chirurgie Générale et Urgences, Avenue du Maréchal de Lattre de Tassigny, 54035 Nancy, France

Abbreviations

ATM line	Asynchronous transfer mode line
ms	Millisecond
PB1	Peg-Board 1
CT2	Camera Targeting 2
TR1	Thread the Ring 1
ED1	Energy Dissection 1

Introduction

Robotic surgery was noted to be in its infancy in 2004, [1] but now this advanced technology is on its way to young adulthood [2]. It has become a standard in complex surgery [3]. The mature experience will likely include the achievement of remote telesurgery, a future challenge for robotic surgeons [4,5].

The first transatlantic human telesurgery procedure was performed in 2001 [6]. Since the proof of concept, telesurgery remains a complex and uncommon process that holds promise in overcoming challenging situations (remote medicine for underserved regions, surgery in the battlefield, surgery in space, etc.) [7,8]. Many teams have worked on the telesurgery process and tried to achieve remote telesurgery procedures using available technical resources for the video flux transfer [7,9,10]. In telesurgery, the control signal sent from the master console is transferred over a network to the robot arms followed by a corresponding movement of the surgical instruments. The video images are then returned to the surgeon site. The data transmission requires an encoding, transmission, and decoding process in which a time delay, or latency, is inevitably produced. Latency is correlated with the amount of data and the quality of network. The first transatlantic human telesurgery (with the Zeus robot) used sophisticated dedicated asynchronous transfer mode (ATM) lines with a transmission delay around 150 ms [6]. Dedicated lines, however, are not always feasible in routine clinical situations. The public Internet bridging the world could be an easy and accessible resource to transmit this data. Even so, the network availability would be at the price of increasing latency measured approximately 450–900 ms [11].

It would be valuable to clarify the impact of the latency on surgical performances before future implementations of telesurgery. Two thresholds need to be established: The first is the smallest latency that can be detected by surgeons which will influence their performance, and the second is the level of latency that makes the surgery unsafe. Unsafe surgery is associated with an increase in errors. A previous study on this topic highlighted the impact of delay on performance degradation using the dV-Trainer®. The authors evaluated the effects of delay varying between 100 and 1000 ms, and found that latencies $\leq$300 ms had a small impact on performance. Subjective evaluation then suggested that surgery became quite difficult at delays $\geq$800 ms [12]. However, this study only included medical students as the subjects. Additional experiments should be performed with experienced surgeons, especially those experienced with robotic systems which would be needed to implement telesurgical procedures.

The present study aims to evaluate, on a surgeon population, the impact of different latency levels on performances in four simulated robotic tasks.

Material and methods

Exercises and subjects

We designed a prospective, observational study conducted on the robotic surgical simulator dV-Trainer® (Mimic technologies Inc., Seattle, USA). This tool has demonstrated face, content, construct, and concurrent validity in previous studies [13,14]. Based on expert opinion and literature review [14,15], we chose four exercises for the test that would be performed in a constant easy-to-difficult order: (a) Peg-Board 1 (PB1)—pick up and transfer rings sequentially from the Peg-Board to a single peg on the floor; (b) Camera Targeting 2 (CT2)—manipulate the camera to precisely focus and zoom on a target sphere; pick up and move a stone into a designated basket; (c) Thread the Rings 1 (TR1)—pass a needle and suture through a number of flexible eyelets; (d) Energy Dissection 1 (ED1)—isolate a large blood vessel by cauterizing and cutting small branching blood vessels that anchor the large vessel (Fig. 1). Both basic (endowrist manipulation, camera control, clutching) and challenging (suturing, dissection) skills were covered with these exercises. The dV-Trainer® simulator permitted us to introduce fixed latencies into the exercises between the gesture on the grips and the visual feedback on the console.

After institutional review board approval, we recruited subjects—fellows and attending surgeons—during three robotic surgery conferences. All the experiments involving human participants were in accordance with the ethical standards of the institutional research committee, as well as the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants.

Procedures

Each participant received a unique identification number under which all his/her data would be collected, and then completed a questionnaire concerning demographic data (including surgical experience and related activities).

Each subject was randomly and blindly assigned a latency varying between 100 and 1000 ms with increments of 100 ms. Before the trials on dV-Trainer®, they received standard instruction on its use in a familiarization period. After that, they performed all four exercises in order without delay (Base). The results provided their baseline performance. Then they repeated the same set of exercises twice with the assigned latency (Warm-up and Test). The Warm-up period allowed them to become familiar with latency and to acquire short-term adaptation (Fig. 2).

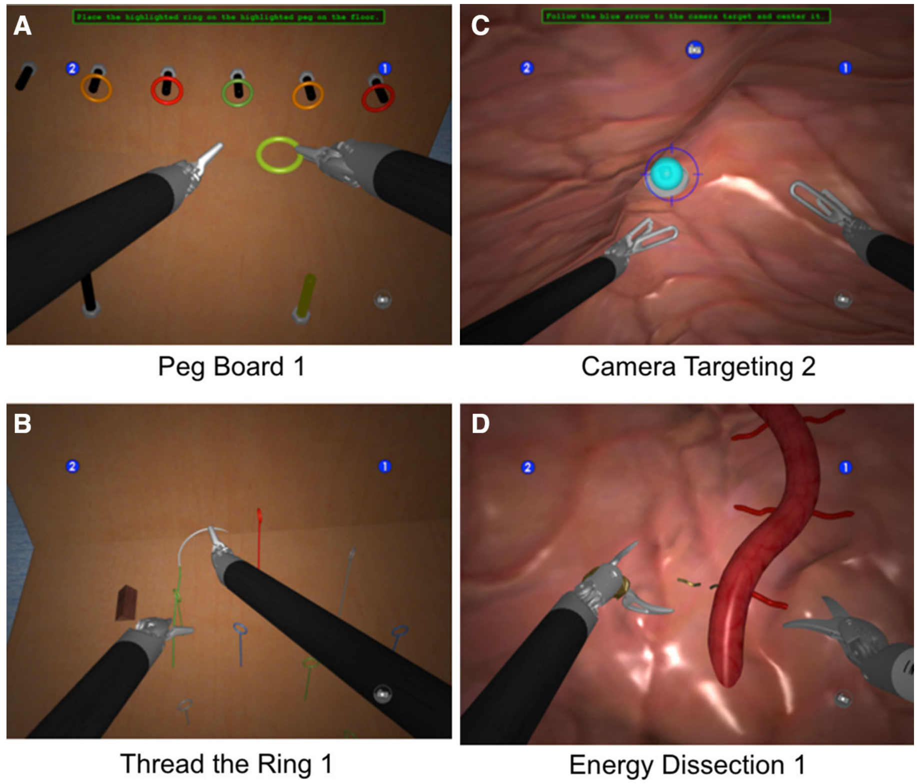


Fig. 1 The four dV-Trainer® exercises: Peg-Board 1 (**a**), Camera Targeting 2 (**b**), Thread the Ring 1 (**c**), and Energy Dissection 1 (**d**)

Assignment of a delay (varying between 100 and 1,000 ms)
⇩
Standard instruction on the use of dV-Trainer®
⇩
Base: one attempt of PB1, CT2, TR1, and ED1 without delay
⇩
Warm-up: one attempt of the four exercises with the assigned delay
⇩
Test: one attempt of the four exercises with the assigned delay

Fig. 2 Experimental procedures

Metrics

The dV-trainer includes a built-in scoring system. The values of the following metrics were automatically recorded after each exercise: time to complete the exercise (in seconds), instrument motion (in centimeters), master workspace range (in centimeters), excessive instrument force (in seconds), instruments out of view (in centimeters), instrument collisions, drops, etc. An overall score representing a combination of these criteria was also automatically generated.

Based on our experience, the task completion time is the most sensitive and reliable measure to the impact of delay [12]. We thus chose this measure to represent the results. In addition, the mean score of all error metrics was calculated in order to evaluate the latency impact on errors.

Statistics

Data were analyzed using the R statistical software. A repeated-measures ANOVA (mixed-effects model) was used to determine the differences in performances between various latency groups (with FDR p value correction), and also between the three periods in each latency group (with Holm correction). Statistical significance was determined at $p < 0.05$.

Results

Complete data

Final data were derived from 37 surgeons. Twenty-three persons had robotic experience, with an average of 2.7 years (ranging from 1 to 9 years). All subjects completed the three stages from Base to Test, but some of them did not complete all the exercises. For example, four subjects were included in the 100 ms group, but one of them did not complete the exercises of CT2 and TR1. The groups from 700 to 1000 ms were combined due to the limited subject number (Table 1).

Results across exercises

The different latency groups achieved similar baseline performance with a mean task completion time of 207.2 s ($p >$

Table 1 Demographic data

	100 ms	200 ms	300 ms	400 ms	500 ms	600 ms	700–1000 ms
n	4	8	2	7	4	7	5
Complete PB1	4	7	2	5	4	6	3
Complete CT2	3	6	2	5	4	4	5
Complete TR1	3	7	2	6	4	1[a]	1[a]
Complete ED1	4	7	2	7	3	3	1[a]
Age (years)	38.5 ± 7.0	45.4±10.5	48.5±13.4	47.4±12.5	37.8±4.3	44.6±8.4	42.6±9.4
Position (*n*)	Fellow (2) attending (2)	Fellow (3) attending (5)	Fellow (0) attending (2)	Fellow (1) attending (6)	Fellow (1) attending (3)	Fellow (0) attending (7)	Fellow (1) attending (4)
Laparoscopic experience (years)	6.3±1.9	11.8±7.9	10.5±3.5	16.7±13.1	7.3±1.5	13.6±6.6	14.0±8.4
Robotic experience (years, median)	0	0	1.25	1	0.5	2	3

[a] Data were not used due to the single number

0.05). An increasing tendency of the task duration with delay was observed in the two latency stages. In the Test period, the mean task duration increased from 156.4 s at 100 ms to 310.7 s at 500 ms. When comparing this measure between any two latency groups, statistical significance was achieved in the comparisons of the 100 ms group versus the 400 and 500 ms groups ($p < 0.05$; Fig. 3).

Subjects demonstrated the tendency to improve their performances from the Warm-up to the Test period. The task completion time deteriorated from the baseline to the two latency stages at 300 ms and above, although statistical significance was not achieved at 300 ms due to the limited subject number (Fig. 3). The comparison results between the three periods in each latency group are illustrated in Fig. 4.

The mean error score deteriorated from baseline to latency stages at 500 ms and above ($p < 0.05$). For example, in

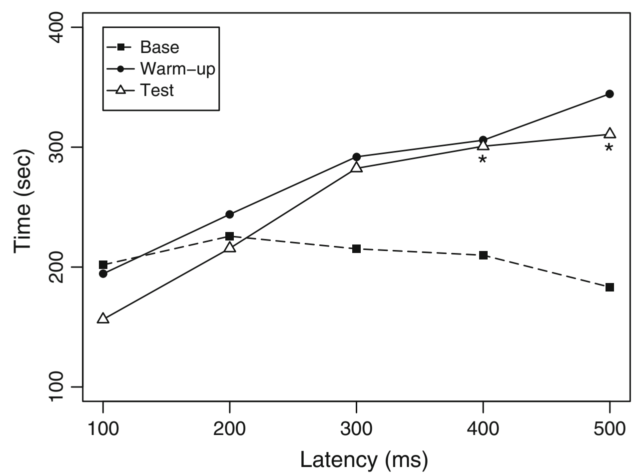


Fig. 3 The mean task completion time across the four test exercises in each latency group. *Difference was determined compared to the 100 ms group ($p < 0.05$). The groups of 600 and 700–1000 ms were not included due to insufficient data in certain exercises

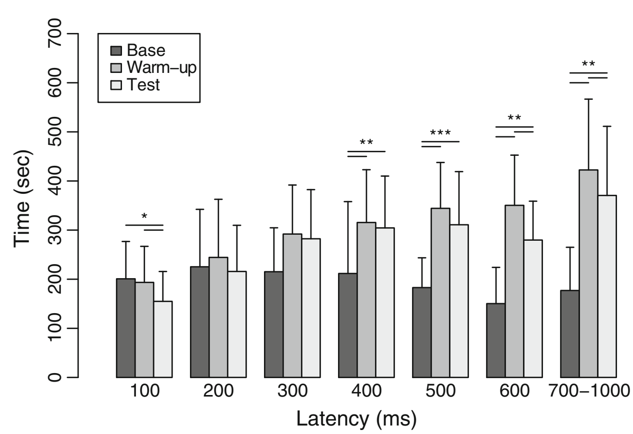


Fig. 4 Comparisons of the task completion time between the three periods in each latency group (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$). The group 600 ms includes only the results across PB1, CT2, and ED1; the group 700–1000 ms includes the results across PB1 and CT2

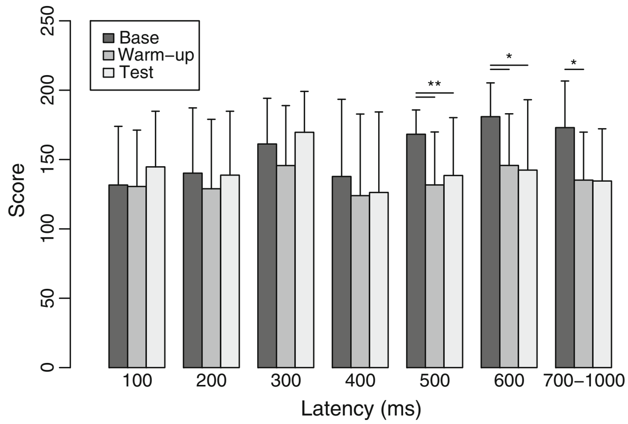


Fig. 5 Comparisons of the mean error score between the three periods in each latency group ($*p < 0.05$; $**p < 0.01$). The group 600 ms includes only the results across PB1, CT2, and ED1; the group 700–1000 ms includes the results across PB1 and CT2

500 ms group, the score decreased from 168.2 (out of 200) to 138.5 from the Base to the Test period (Fig. 5).

Results in separate exercises

An increasing tendency of the task completion time with latency was observed in the two latency periods of the four exercises. The degradation of performances between baseline and latency stages started with 300, 500, 100, and 300 ms in PB1, CT2, TR1, and ED1, respectively (Fig. 6).

Incomplete data

Eighty incomplete exercises in latency stages derived from 26 subjects were identified. They included 18 PB1, 18 CT2, 26 TR1, and 18 ED1. Subjects were physically unable to complete these delayed exercises. Fifty-three (66.25 %) exercises were stopped by the subjects at a mean time of 9.8 min (586.01 ± 14.54 s). The ratio of incomplete exercises was relatively higher in high-delay groups (Fig. 7).

Discussion

We aimed to determine the latency effects on surgical performances in experienced surgeons who are unfamiliar with latency and the simulator device, to establish the threshold delays in telesurgery. Overall, the gradually increasing latency has an increasing impact on performances, and the performance deterioration consistently begins at 300ms. Latencies of 100 and 200 ms seemed to have no clear effect, and the 100 ms group had improving performance from the Base to the Test stage. This improvement likely corresponds to the learning effects of basic simulator manipulation and

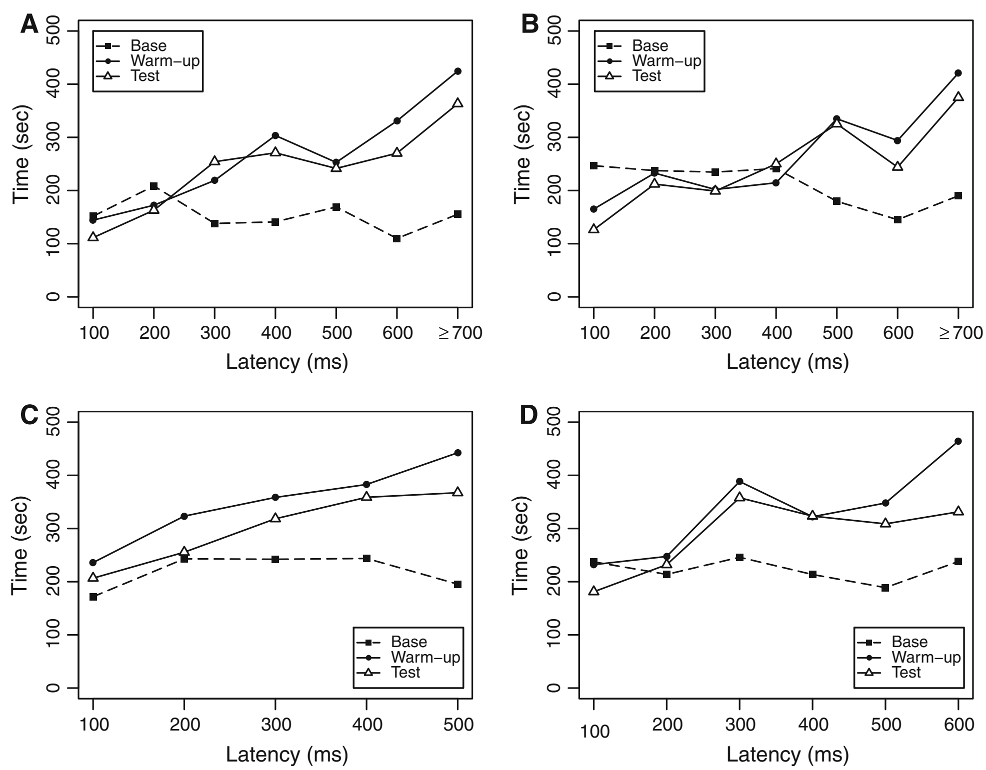


Fig. 6 The mean task completion time in each latency group of the four exercises: Peg-Board 1 (**a**), Camera Targeting 2 (**b**), Thread the Ring 1 (**c**), and Energy Dissection 1 (**d**)

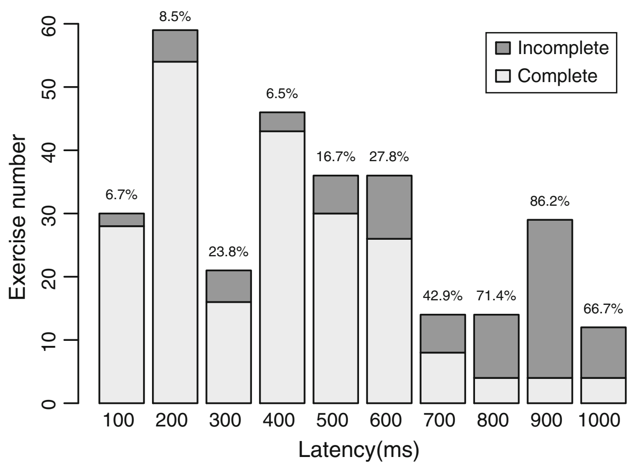


Fig. 7 The numbers of complete and identified incomplete exercises at each latency level. The numbers above the *bars* represent the percentage of incomplete exercises

further proves that 100 ms does not have a significant influence. For the superior threshold, delays equal or higher than 700 ms seem to be difficult to manage especially in complex tasks. Only one subject was able to complete the tasks at 700 ms, and only the easiest exercises (PB1, CT2) were finished at 800–1000 m. In the previous study with trained medical students, the similar threshold was highlighted and the authors suggested telementoring as a safer choice [12]. Telementoring is an application of telemedicine that involves the remote guidance of a procedure when the local operator has limited experience with the technique [16]. However, in this study, the error rate significantly increased from non-latency to latency stages at delays ≥500 ms, which may indicate an increase in surgical risk. We would consider this value as the superior threshold, and telesurgery should not be recommended in this condition for most surgeons [17]. This does not mean that procedures cannot be performed at higher-latency levels, and results could be better for experienced robotic surgeons, especially when given an opportunity to rehearse in an environment including latency, such as with a simulator. Current research is still limited, and outcome data are lacking to demonstrate the feasibility and safety of telesurgery with high delays. In a previous published study, a nephrectomy was performed on a swine under a delay of 900 ms. Two surgeons performed the procedure, one in the remote site console and the other in the local site console [11]. In this article, no outcome data were provided, such as surgical performance and the mental stress of surgeons.

Surgeons have been shown to have the potentials to adapt to delays [18]. Similar tendency was also observed in our study: Performances improved from Warm-up to Test. It suggests that surgeons may be trained on latency to improve their telesurgical performance. However, the improvement observed here is not clearly attributable to adaptation through experience with latency. It may also be the result of improvements in psychomotor simulator manipulation. Despite the overall tendency across exercises, our results also demonstrate that the impact of latency is related to the difficulty of procedures. Latency affected performances on different levels for the four chosen exercises: The performance deterioration started at a high delay (500 ms) for the simple exercise CT2 and at a low level (100 ms) in the more challenging TR1. This fact indicates that the minimum influential and the maximum acceptable delays could be different in surgical procedures with different complexity.

For the challenging exercises that may better represent real surgical scenarios, we have chosen TR1 and ED1 instead of the more complex exercises like "Suture Sponge" or "Tubes." This is because many surgeons were not sufficiently familiar (or proficient) with the robot or the simulator. In this study, few tasks were completed at delays higher than 700 ms. One might anticipate that the results would be even worse if applying more challenging exercises.

Participants have demonstrated the efforts to complete the tasks even with considerable latencies. In the identified 80 incomplete exercises, only a few subjects terminated their participation soon after beginning. The mean duration of attempt was 7.5 min per exercise. This effort could minimize the bias of experiments. It is also interesting to observe that many persons stopped at about 10 min. It seems that this is a threshold beyond which surgeons could no longer endure the effects of latency.

This study has potential limitations: Although we recruited more than 60 surgeons, the final completion rate was lower than expected. The small number of subjects in each latency group is a shortcoming of the study. We did not merge different latency groups because the objective was to evaluate the impact of each latency level, and an interval of 100 ms may already cause difference. Also the distribution of subjects was not equivalent in different latency groups, primarily due to subjects choosing to terminate their participation before completing the entire experiment. Fewer subjects were included in the 300 ms group. Moreover, many surgeons failed to complete the tasks at high delays due to the difficulty of manipulation under these conditions. In addition, all subjects were novices in telesurgery (or latencies) since this technology is currently only available in research settings.

A complementary study will be necessary to assess the performance degradation induced by latency on robotic surgery experts, and to investigate whether latency training could be used to overcome the challenges of telesurgery.

Conclusion

This study was conducted on surgeons with limited experience using the dV-Trainer simulator, and the results demon-

strated that performances (time to perform, score, error) deteriorate gradually as latency increases. The impact of delay is related to the difficulty of the procedures, but overall, delays of 100 to 200 ms have no significant impact, and a delay higher than 500 ms causes a noticeable increase in surgical risk. Surgery becomes extremely difficult and should be avoided at delays higher than 700 ms. Telementoring could be an option in this situation. Surgeons have the potential to adapt to latency, and they may be trained to improve their telesurgical performances using devices like simulators of robotic systems.

Compliance with ethical standard

Conflict of interest The authors Manuela Perez, Song Xu, Sanket Chauhan, Alyssa Tanaka, Khara Simpson, Haidar Abdul-Muhsin and Roger Smith declare they have no disclosure or conflict of interest to declare

References

1. Lanfranco AR, Castellanos AE, Desai JP, Meyers WC (2004) Robotic surgery: a current perspective. Ann Surg 239:14–21. doi:10.1097/01.sla.0000103020.19595.7d
2. Kim SP, Boorjian SA, Shah ND, Weight CJ, Tilburt JC, Han LC, Thompson RH, Trinh QD, Sun M, Moriarty JP, Karnes RJ (2013) Disparities in access to hospitals with robotic surgery for patients with prostate cancer undergoing radical prostatectomy. J Urol 189:514–520. doi:10.1016/j.juro.2012.09.033
3. Ghani KR, Sukumar S, Sammon JD, Rogers CG, Trinh QD, Menon M (2014) Practice patterns and outcomes of open and minimally invasive partial nephrectomy since the introduction of robotic partial nephrectomy: results from the nationwide inpatient sample. J Urol 191:907–912. doi:10.1016/j.juro.2013.10.099
4. Haidegger T, Sándor J, Benyó Z (2011) Surgery in space: the future of robotic telesurgery. Surg Endosc 25:681–690. doi:10.1007/s00464-010-1243-3
5. Gambadauro P, Torrejón R (2013) The "tele" factor in surgery today and tomorrow: implications for surgical training and education. Surg Today 43:115–122. doi:10.1007/s00595-012-0267-9
6. Marescaux J, Leroy J, Gagner M, Rubino F, Mutter D, Vix M, Butner SE, Smith MK (2001) Transatlantic robot-assisted telesurgery. Nature 413:379–380. doi:10.1038/35096636
7. Anvari M, McKinley C, Stein H (2005) Establishment of the world's first telerobotic remote surgical service: for provision of advanced laparoscopic surgery in a rural community. Ann Surg 241:460–464
8. Anvari M (2007) Telesurgery: remote knowledge translation in clinical surgery. World J Surg 31:1545–1550. doi:10.1007/s00268-007-9076-5
9. Rayman R, Croome K, Galbraith N, CcCulure R, Morady R, Peterson S, Smith S, Subotic V, Van Wynsberghe A, Primak S (2006) Long-distance robotic telesurgery: a feasibility study for care in remote environments. Int J Med Robot Comput Assist Surg MRCAS 2:216–224. doi:10.1002/rcs.99
10. Rayman R, Croome K, Galbraith N, CcCulure R, Morady R, Peterson S, Smith S, Subotic V, Van Wynsberghe A, Pate R, Primak S (2007) Robotic telesurgery: a real-world comparison of ground- and satellite-based internet performance. Int J Med Robot 3:111–116. doi:10.1002/rcs.133
11. Sterbis JR, Hanly EJ, Herman BC, Marohn MR, Broderick TJ, Shih SP, Harnett B, Doarn C, Schenkman NS (2008) Transcontinental telesurgical nephrectomy using the da Vinci robot in a porcine model. Urology 71:971–973. doi:10.1016/j.urology.2007.11.027
12. Xu S, Perez M, Yang K, Perrenot C, Felblinger J, Hubert J (2014) Determination of the latency effects on surgical performance and the acceptable latency levels in telesurgery using the dV-Trainer$^{(\circledR)}$ simulator. Surg Endosc. doi:10.1007/s00464-014-3504-z
13. Lee JY, Mucksavage P, Kerbl DC, Huynh VB, EtafyM Mc, Dougall EM (2012) Validation study of a virtual reality robotic simulator-role as an assessment tool? J Urol 187:998–1002. doi:10.1016/j.juro.2011.10.160
14. Perrenot C, Perez M, Tran N, Jehl JP, Felblinger J, Bresler L, Hubert J (2012) The virtual reality simulator dV-Trainer$^{(\circledR)}$ is a valid assessment tool for robotic surgical skills. Surg Endosc 26:2587–2593. doi:10.1007/s00464-012-2237-0
15. Ruparel RK, Taylor AS, Patel J, Patel V, Heckman MG, Rawal B, Leveillee RJ, Thiel DD (2014) Assessment of virtual reality robotic simulation performance by urology resident trainees. J Surg Educ 71:302–308. doi:10.1016/j.jsurg.2013.09.009
16. Rosser JC Jr, Gabriel N, Herman B, Murayama M (2001) Telementoring and teleproctoring. World J Surg 25:1438–1448
17. Lendvay TS, Hannaford B, Satava RM (2013) Future of robotic surgery. Cancer J Sudbury Mass 19:109–119. doi:10.1097/PPO.0b013e31828bf822
18. Nguan CY, Morady R, Wang C, Harrison D, Browning D, Rayman R, Luke PPW (2008) Robotic pyeloplasty using internet protocol and satellite network-based telesurgery. Int J Med Robot 4:10–14. doi:10.1002/rcs.173

The International Journal of **Medical Robotics** and **Computer Assisted Surgery**

THE INTERNATIONAL JOURNAL OF MEDICAL ROBOTICS AND COMPUTER ASSISTED SURGERY
Int J Med Robotics Comput Assist Surg 2014; **10**: 379–384.
Published online 26 November 2013 in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/rcs.1559

ORIGINAL ARTICLE

Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development

Roger Smith[1*]
Vipul Patel[2]
Richard Satava[3]

[1]*Florida Hospital Nicholson Center, Celebration, FL, USA*

[2]*Global Robotics Institute, Celebration Health, Celebration, FL, USA*

[3]*University of Washington College of Medicine, Seattle, WA, USA*

*Correspondence to: R. Smith, Florida Hospital Nicholson Center, 404 Celebration Place, Celebration, FL 34747, USA.
E-mail: Roger.smith@flhosp.org

Accepted: 23 October 2013

Abstract

Background There is a need for a standardized curriculum for training and assessment of robotic surgeons to proficiency, followed by high-stakes testing (HST) for certification.

Methods To standardize the curriculum and certification of robotic surgeons, a series of consensus conferences attended by 14 leading international surgical societies have been used to compile the outcomes measures and curriculum that should form the basis for a Fundamentals of Robotic Surgery (FRS) programme.

Results A set of 25 outcomes measures and a curriculum for teaching the skills needed to safely use current generation surgical robotic systems has been developed and accepted by a committee of experienced robotic surgeons across 14 specialties.

Conclusions A standardized process for certifying the skills of a robotic surgeon has begun to emerge. The work described here documents both the processes used for developing educational material and the educational content of a robotic curriculum. Copyright © 2013 John Wiley & Sons, Ltd.

Keywords robotic surgery; outcomes measures; educational curriculum

Introduction

In 2004, the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) launched the validated Fundamentals of Laparoscopic Surgery (FLS) curriculum and, together with the American College of Surgeons (ACS), promoted the FLS as a minimum standard before a surgeon should be allowed to perform laparoscopic procedures independently (1). In 2009, the American Board of Surgery (ABS) mandated that, in addition to Advanced Cardiac Life Support (ACLS) and Advanced Trauma Life Support (ATLS), a certificate documenting the successful passing of the FLS exam be included in the application in order to be eligible to sit the examination for certification in General Surgery (2).

During the last decade, robotic surgery has grown through a similar evolution to laparoscopic surgery and is being recognized as an important surgical approach by multiple surgical specialties. Furthermore, it shows every sign of continuing the adoption of more diverse surgical procedures, as manifested by

the fact that in calendar year 2012, approximately 450,000 robotic surgical procedures were performed. The number of procedures being performed by robotic surgery has been constantly rising in urology, gynaecology, colorectal, paediatric and numerous other specialties. Expert robotic surgeons and numerous surgical societies and certifying organizations have advocated the need for a unified approach and standardized curriculum for basic training, assessment, testing and certification in robotic surgery skills (3). There have been previous efforts to develop a core curriculum for certifying robotic surgeons (4,5); however, these have been fragmented, with different approaches and outcomes measures emerging from each. This has resulted in conflicting, competing and redundant curricula for the training and the assessment tools for robotic surgery. In addition, these curricula have generally lacked the human and financial resources necessary to complete the most comprehensive, multi-institutional validation that is necessary to gain acceptance at a national level.

Through the combined support of two grants, one to the Minimally Invasive Robotics Association and the other to the Florida Hospital Nicholson Center, a process has been created by a multi-specialty group of participants which unifies the previous attempts to develop a robotic curriculum, which included the current developers of the existing curricula and which expand into a much larger foundation of surgical societies with a stake in this new technology. These grants provide the necessary funding to carry the effort through multi-institutional validation with the support of participants who represent all surgical specialties that are currently performing robotic surgery.

The scope of the curriculum development was limited to the creation of a curriculum (course) that encompassed the cognitive, psychomotor and team training skills required to safely operate a surgical robotic system and perform the most basic of manual and communication skills that would be needed to safely perform any robotic surgery procedure. This curriculum does not include the needs assessment, gap analysis, pre-operative care or post-operative care – it is a skills-based curriculum focused upon the skills needed in the operating room. It is assumed that all the above pre- and post-operative education and training has been completed before bringing the patient into the operating room.

Materials and methods

Participation in this effort was invited from multiple certifying boards, professional surgical societies and associations that represent international practitioners and regulators of various surgical specialties, as well as the US Department of Defense (DoD) and Veterans Health Administration (VHA). The conference participants were official representative members of these organizations or agencies and were selected by their organizations to provide insight into the needs of the organization. A complete list of the participating organizations is given in Table 1 and a list of the individual participants is given in the Acknowledgements section of this paper. While they do not formally represent an endorsement or acceptance of the results at this interim period of the curriculum development, and their participation does not imply acceptance by the societies, boards or agencies, at the completion of the validation trials, the organizations will review the final results for endorsement or as a requirement for surgical training. This project is an effort to provide the stakeholders with the best scientific evidence upon which to base their decisions regarding implementation of a fundamental curriculum to meet their needs, while reducing redundancy, competition and duplication of effort.

Each consensus conference was conducted over a 2 day period, using a task analysis followed by a modified Delphi method (6) to achieve consensus on the materials that were created and accepted by the group. The concepts and criteria contributed by the members were analysed for commonality to create a list of critical items in robotic surgery. Previously published material from a single institution's curriculum was used as a template for initial idea generation (7,8). The individual outcomes measures and curriculum materials were itemized and votes taken on their importance according to each participant. This method led to a composite ranking of outcomes measures which was captured in a draft report. This report was then circulated to each participant for his/her private, anonymous deliberation (classic Delphi method). Following the editing of the comments on the initial draft, a second classic Delphi round was sent to each participant, who then submitted a second set of scores, which were informed by the first composite scores but anonymous to other group members. This classic Delphi method led to a higher level of consensus around the measures and the curriculum. It also identified those items for which there was little group support. Items with

Table 1. Organizational representation in fundamentals of robotic surgery

Accreditation Council of Graduate Medical Education (ACGME)
American Association Gynecologic Laparoscopy (AAGL)*
American College of Surgeons (ACS)
American Congress of Obstetrics and-Gynecology (ACOG)
American Academy of Orthopedic Surgeons (AAOA)
American Association of Colo-rectal Surgeons (ASCRS)
American Association of Thoracic Surgeons (AATS)
American Board of Surgery (ABS)
American Urologic Association (AUA)*
Association of Surgical Educators (ASE)
European Urology Association (EUA)
Minimally Invasive Robotic Association (MIRA)†
Society of American Gastrointestinal and Endoscopic Surgeons (SAGES)*
Society for Robotic Surgery (SRS)
Residency Review Committee (RRC) – Surgery
Royal College of Surgeons-Australia (RCSA)
Royal College of Surgeons-Ireland (RCSI)
Royal College of Surgeons-London (RCSL)
US Department of Defense (DoD)†
US Department of Veterans Health Affairs (VHA)

*Official representative participation.
†Funding organizations.

DOI: 10.1002/rcs

little group support were removed from the list of outcomes measures and from the outline of the curriculum.

The first conference on outcomes measures was attended by 20 participants, including surgeons, scientists, educators, representatives of governing and certification organizations and facilitators. The ranking of the tasks identified was done by a subset of nine experienced clinical surgeons. Participants who were not surgeons abstained from the scoring process.

The second conference on curriculum development was attended by 38 surgeons, scientists, medical educators, behavioural psychologists, psychometricians and facilitators. This group reviewed and became familiar with the material from the first conference. Thereupon, they were divided into three working groups to develop the detailed information in the curriculum that focused on didactic and knowledge-based information, psychomotor skills and team training and communications. At the conclusion of the three focused workshops, all participants reviewed the report of the separate workgroups, consolidated the three sections back into a single curriculum, which was then deliberated until consensus was reached and then voted upon. Similarly, the final ranking of the material developed was limited to experienced surgeons within the group. The role of the scientists, educators, psychologists and psychometricians was to ensure that the material created by the surgeons was structured into effective and valid educational and testing products.

The products from these meetings will go through a multi-site validation trial in which subjects are trained using these materials and the results collected, evaluated and used to modify the materials as necessary.

Results

The first consensus conference (outcomes measures) resulted in a list of 25 outcomes measures, which the group agreed should be the minimal skills needed by a surgeon seeking to safely perform robotic surgery, regardless of his/her specialty. These included eight pre-operative, 15 intra-operative and two post-operative skills, which are shown in Table 2. The resulting documents also provided detailed definitions, descriptions, errors, outcomes and metrics for each of these skills [Martino M. Basic skills curriculum for robotic gynecologic surgery (unpublished)].

The second and third consensus conferences (curriculum development), which focused on actually creating the curriculum and its content, initially resulted in outlines and principles for the creation of a curriculum to teach the previously identified list of skills and knowledge (Table 3) (9). This document was then expanded into a fully detailed curriculum by clinical surgeons working in conjunction with experienced surgical educators, behavioural psychologists, statisticians and psychometicians. The result was a full life-cycle curriculum that consists of three components: cognitive skills, psychomotor skills, and team training and communication skills.

Table 2. FRS outcomes measures

Pre-operative	Intra-operative	Post-operative
System settings	Energy sources	Transition to bedside assistant
Ergonomic positioning	Camera control	Undocking
Docking	Clutching	
Robotic trocars	Instrument exchange	
Operating room set-up	Foreign body management	
Situation awareness	Multi-arm control	
Closed-loop communicationss	Eye–hand instrument coordination	
Response to system errors	Wrist articulation	
	Atraumatic tissue handling	
	Dissection – fine and blunt cutting	
	Needle driving	
	Suture handling	
	Knot tying	
	Safety of operative field	

Table 3. FRS curriculum principles

Cognitive	Psychomotor skills	Team training
Lecture and video	Physical test device	Interdisciplinary team
Introduction to robotic systems	Single integrated device	WHO pre-operative checklist
Pre-operative activity	3D working space	Robotic-specific communication
Intra-operative activity	Based on existing validated tasks	Post-operative debriefing
Post-operative activity	Affordable design	Team crisis response
Each activity includes goals, conditions, standards, metrics, errors	High fidelity for examination, lower fidelity for training	
	Ease and reliability of scoring	

DOI: 10.1002/rcs

Cognitive skills

The didactic and cognitive (knowledge base) working group created an outline of the material which should be taught in lecture format. Since the training was in basic skills (not surgical procedures), there were no 'steps of the procedure' which are traditionally included when developing a procedure-focused curriculum. This curriculum included:

1. Introduction to the principles and functionality of robotic surgical devices.
2. Pre-operative set-up of equipment and positioning of patient and staff, placement of ports, check lists and all activities required of the surgeon before sitting at the robotic console.
3. Intra-operative use of a robot, description of the critical psychomotor skills, surgeon ergonomics, visual field control, operative control of the robot, necessary instruments and supplies. Also included were surgeon communication skills from the console to the operating room team (team training).
4. Post-operative steps for shutting down the robot, removing a robot from the operative field and transitioning the patient to a gurney.

Each of these included an explicit list of passing criteria and errors that can occur in the process.

Psychomotor skills

The psychomotor skills working group initiated their work by defining the seven principles that should be applied in selecting or designing a psychomotor skills device for robotic surgery. Those principles were:

1. The tasks should be three-dimensional in nature.
2. The tasks designed for testing should be such that they have multiple learning objectives that incorporate multiple skills from the outcomes measures.
3. The skills should be designed to train the full capability of the robotic system, to include skills and tasks that are not possible in open or laparoscopic surgery.
4. Implementation of the tasks and the resultant method for teaching should not be cost-prohibitive.
5. High-fidelity models should be used for testing. Training can use lower-fidelity devices and methods.
6. Tasks should be easy to administer to ensure inter-rater reliability (IRR).
7. The tasks should be designed for implementation with physical objects and devices. The device will be developed initially in virtual reality (VR) as a CAD/CAM model, from which the actual physical models will be 'printed' with stereolithography (the VR model objects will be identical to the physical objects), creating a training experience that would be identical in both the virtual and real world.

The group then identified 16 of the 25 skills that contained psychomotor features. In order to implement this psychomotor skills curriculum, 10 tasks were created for the dome-shaped device, which could be used to train and measure the 16 skills. Three tasks were drawn from FLS (with slight modifications); others were selected from existing educational programmes presented by participants and found in the published literature (4,7,8), and designs for new task devices were proposed and debated by the participating surgeons:

1. FLS peg transfer.
2. FLS suturing and knot tying.
3. FLS pattern cutting.
4. Running suture.
5. Dome with four towers for ambidexterity.
6. Vessel dissection and clipping.
7. Fourth arm retraction and cutting.
8. Energy and mechanical cutting.
9. Docking task (new design).
10. Trocar insertion task (new design).

For each of these, the group also identified the associated task description, conditions, metrics and errors. These details can be found in the milestone report of the event (10).

The group felt that, for ease and simplicity of implementing the training of the tasks and skills, it was important for all of these tasks to be performed on a single integrated device, which could be scored by visual observation for training, assessment and testing. Incorporated into the planning was that the design of the device would permit future adaptation of more sophisticated automated scoring by an identical device which was either a mechanical or computer-based (VR) version. Toward this end, they created the initial design for the 'FRS dome' shown in Figure 1. Prototypes of this dome have been created to test the usability and reliability of the device itself during a pilot study, and the training and assessment effectiveness of the device will be evaluated during the multi-institutional validation trials of the FRS curriculum.

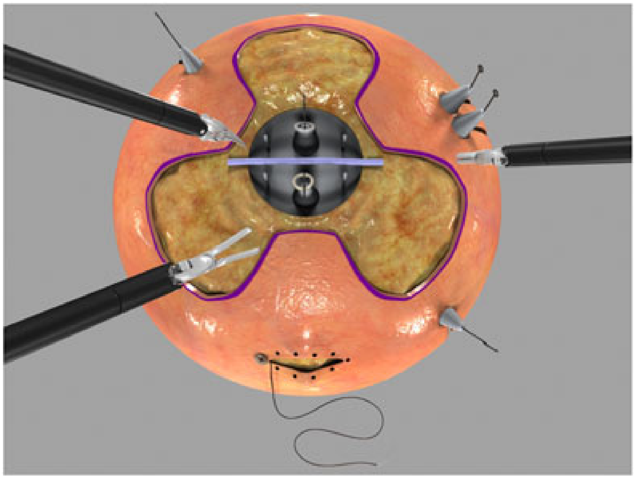


Figure 1. FRS dome device design

DOI: 10.1002/rcs

Team training and communications

The team training and communications working group prefaced their work by defining the importance of team training in a robotic environment. They identified the following principles as essential to successful team-based operations and training:

1. Inclusion.
2. Empowerment.
3. Person-specific.
4. Reiterative.
5. 'Just in time'.
6. Ownership.
7. Risk management/quality improvement – closed loop communication.

They stated that existing programmes, such as TeamSTEPPS®, can be applied to robotic teams. Their curriculum follows a checklist format and is conceptually derived from the standard WHO checklist. For robotic training they recommended the following checklists:

1. *Pre-operative*. Addressing general situation, surgeon, anaesthetist, nurse/OPD, and surgical site infection and robotic docking, in addition to addressing anaesthesia, patient, bedside assist, procedure-specific checks and trouble shooting.
2. *Intra-operative*. Addressing the communication that occurs within a team throughout the operation. Special emphasis was placed upon developing communication skills for the surgeon and the team once surgery has begun, because the surgeon has no visual contact with the remainder of the operating team during the procedure.
3. *Post-operative*. Undocking, patient transport and final debriefing.

Based on these, the groups generated outlines for a full curriculum to teach these information and communication skills. Those were then expanded into a full curriculum by an experienced surgical educator, which is currently being developed into a publicly accessible online education system.

Discussion

A consensus conference process, involving members from major stakeholder organizations in surgical training, governance and certification across multiple specialties, was implemented, with the result of a curriculum for the most important outcome measures for the safe conduct of robotic surgery. The development of FRS is multi-specialty, system agnostic and follows decades of experience in other industries at developing such education, training and assessment platforms.

This curriculum for training and assessment should be executed not by a time-based course limited by number of days, sessions, etc., but rather in a competency-based fashion, that is, continuing to train until the student's learning curve has reached the benchmark values (set by the mean of the learning curve of experienced surgeons). With such training and assessment, a learner should be able to demonstrate proficiency in basic robotic surgery skills and should be capable of passing the requirements of high-stakes testing and evaluation that would lead to his/her certification in technical skills. The current training programme, which has been designed by and is taught by the device manufacturer, presents excellent information and hands-on experience with the equipment. However, it is a time-based exposure and attendance programme, with no measurement of the proficiency of the attendees.

The curriculum, conference reports and associated artifacts from the process will be transitioned to an independent, objective, unbiased professional organization with the mission and capability to develop and administer testing in a manner that meets requirements for certification. The goal of this specific manuscript is to provide a detailed process and methodology for developing a curriculum that is template-based, which is easy to use, flexible to meet the needs of many different specialties, reduces redundancy and competition and, because of its modularity, is cost efficient and reduces the time to develop subsequent curricula.

Acknowledgements

This project is a collaboration of leading robotic surgeons and educators. The following have all participated in, and contributed to, the creation of the materials reported here: A. Advincula; R. Aggarwal; A. Al Ansari; D. Albala; R. Angelo; M. Anvari; J. Armstrong; G. Ballantyne; M. Billia; J. Borin; D. Bouchier-Hayes; T. Brand; J. Cannon-Bowers; S. Chauhan; R. Coelho; S. Cuschieri; B. Dunkin; S. Dunlow; V. Ficarra; A. Gallagher; L. Glazerman; A. Goh; T. Grantcharov; D. Hananel; J. Hebert; R. Holloway; J. Hubert; W. Judd; P. Kenneth; K. Kim; M. Koch; T. Kowalewski; R. Kumar; K. Kunkler; G. Lee; T. Lendvay; R. Leveillee; J. Levy; G. Lewis; G. Maddern; S. Magnuson; P. Manuela; M. Marohn; D. Maron; M. Martino; P. Neary; E. Parra-Davila; V. Patel; C. Perrenot; S. Ramamoorthy; K. Rha; J. Riess; B. Rocco; R. Rush; R. Satava; B. Sayers; D. Scott; N. Seymour; N. Siddiqui; M. Sinanan; R. Smith; D. Stefanidis; C. Sundaram; R. Sweet; E. Verrier; G. Weinstein; V. Wilcox. This study was collaboratively supported by grants to the Minimally Invasive Robotics Association and to the Florida Hospital Nicholson Center, and partially sponsored by the US Department of the Army (Award No. W81XWH-11-2-0158) to the recipient Adventist Health System/Sunbelt Inc. (Florida Hospital Nicholson Center); the US Army Medical Research Acquisition Activity (820 Chandler Street, Fort Detrick, MD 21702–5014, USA) is the awarding and administering acquisition office. The

content of the information does not necessarily reflect the position or the policy of the US Government, and no official endorsement should be inferred.

References

1. Peters JH, Fried GM, Swanstrom LL, *et al*. SAGES FLS Committee. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery* 2004; **135**(1): 21–27.
2. ABS news release. ABS to require ACLS, ATLS and FLS for general surgery certification; 15 August 2008: http://www.absurgery.org/default.jsp?news_newreqs [accessed 6 November 2012].
3. Zorn KC, Gautam G, Shalhav AL, *et al*. Training, credentialing, proctoring and medicolegal risks of robotic urological surgery: recommendations of the Society of Urologic Robotic Surgeons. *J Urol* 2009; **182**(3): 1126–1132.
4. Goh AC, Goldfarb DW, Sander JC, *et al*. Global evaluative assessment of robotic skills: validation of a clinical assessment tool to measure robotic surgical skills. *J Urol* 2012; **187**(1): 247–252.
5. Hung AJ, Patil MB, Zehnder P, *et al*. Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. *J Urol* 2012; **187**(2): 630–637.
6. Dalkey N. The Delphi method: an experimental study of group opinion. RAND Corp. Memorandum (RM-5888-PR), 1969.
7. Dulan G, Rege RV, Hogg DC, *et al*. Proficiency-based training for robotic surgery: construct validity, workload, and expert levels for nine inanimate exercises. *Surg Endosc* 2012; **26**(6): 1516–1521.
8. Dulan G, Rege RV, Hogg DC, *et al*. Content and face validity of a comprehensive robotic skills training program for general surgery, urology, and gynecology. *Am J Surg* 2012; **203**(4): 535–539.
9. Satava RM, Smith RS, Patel VR, et al. Fundamentals of Robotic Surgery Consensus Conference I: Outcomes Measures, report from consensus conference.
10. Satava RM, Smith RS, Patel VR, *et al*. Fundamentals of Robotic Surgery Consensus Conference II: Curriculum Development, report from consensus conference.

DOI: 10.1002/rcs

JOURNAL OF ENDOUROLOGY
Volume 29, Number 10, October 2015
© Mary Ann Liebert, Inc.
Pp. 1183–1188
DOI: 10.1089/end.2015.0104

Crowd-Sourced Assessment of Technical Skills: Differentiating Animate Surgical Skill Through the Wisdom of Crowds

Daniel Holst, BS,[1] Timothy M. Kowalewski, PhD,[2] Lee W. White, PhD,[3] Timothy C. Brand, MD,[4] Jonathan D. Harper, MD,[5] Mathew D. Sorensen, MD,[5] Mireille Truong, MD,[6] Khara Simpson, MD,[6] Alyssa Tanaka, PhD,[6] Roger Smith, PhD,[6] and Thomas S. Lendvay, MD[5]

Abstract

Background: Objective quantification of surgical skill is imperative as we enter a healthcare environment of quality improvement and performance-based reimbursement. The gold standard tools are infrequently used due to time-intensiveness, cost inefficiency, and lack of standard practices. We hypothesized that valid performance scores of surgical skill can be obtained through crowdsourcing.

Methods: Twelve surgeons of varying robotic surgical experience performed live porcine robot-assisted urinary bladder closures. Blinded video-recorded performances were scored by expert surgeon graders and by Amazon's Mechanical Turk crowdsourcing crowd workers using the Global Evaluative Assessment of Robotic Skills tool assessing five technical skills domains. Seven expert graders and 50 unique Mechanical Turkers (each paid $0.75/survey) evaluated each video. Global assessment scores were analyzed for correlation and agreement.

Results: Six hundred Mechanical Turkers completed the surveys in less than 5 hours, while seven surgeon graders took 14 days. The duration of video clips ranged from 2 to 11 minutes. The correlation coefficient between the Turkers' and expert graders' scores was 0.95 and Cronbach's Alpha was 0.93. Inter-rater reliability among the surgeon graders was 0.89.

Conclusion: Crowdsourcing surgical skills assessment yielded rapid inexpensive agreement with global performance scores given by expert surgeon graders. The crowdsourcing method may provide surgical educators and medical institutions with a boundless number of procedural skills assessors to efficiently quantify technical skills for use in trainee advancement and hospital quality improvement.

Introduction

The healthcare environment is shifting toward performance-based reimbursement and focusing on quality improvement. A 2000 study from the Agency for Healthcare Research and Quality showed that the surgical mortality rate is among the top 10 causes of death in the United States.[1] While not all deaths from surgery were due to technical errors in this particular report, a different study, which focused on the role of surgical trainees, showed that 56% of malpractice claims unearthed errors in the manual technique.[2]

Recent literature has shown that blinded video assessments of technical performances among experienced laparoscopic surgeons directly correlate with patient outcomes.[3] Subsequently, efforts have been made to adopt methods for evaluating technical skill with tools such as GEARS (Global Evaluative Assessment of Robotic Skills) and GOALS (Global Objective Assessment of Laparoscopic Skills). Both are surgical performance scales that have been extensively validated for use in grading surgical technical skill.[4,5] They are gold standard methods for evaluating surgical performances objectively, but are often burdensome and require too much time and too many resources, yielding these methods impractical for frequent use. In addition, scaling these methods to much larger studies is not practical and, in many cases, not possible.

[1]University of Washington School of Medicine, Seattle, Washington.
[2]Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minnesota.
[3]Stanford University School of Medicine, Palo Alto, California.
[4]Department of Urology, Madigan Army Medical Center, Tacoma.
[5]Department of Urology, University of Washington, Seattle, Washington.
[6]Florida Hospital Nicholson Center, Orlando, Florida.

Crowdsourcing is the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people, especially from the online community rather than from traditional employees or suppliers.[6] The advent of the Internet has enabled the global labor market ready to perform various tasks/surveys to help solve problems. These problems differ widely in scope, yet crowdsourcing is a common denominator used in helping to solve them. Examples include an app used to help solve protein-folding problems and another to help blind users find their mobile phone.[7,8] In recent studies by Chen et al. and Holst et al., crowds have been shown to be as effective as expert surgeons at evaluating surgical technical skill in a dry-laboratory setting.[9,10] Not only did the crowds perform as effectively as the expert surgeons in providing skill assessment but also the cost efficiency and practicality of use were all improved with crowd graders compared to expert surgeon graders. The major limitation of these studies was that the surgical tasks being assessed were dry-laboratory tasks. Thus, no real tissue was being manipulated in the study, leaving questions regarding whether nonexperts can appreciate the subtlety of real surgery. In this study, we hypothesize that crowdsourcing can be used to obtain valid performance grading of surgical technical skill on real, living viable tissue.

Materials and Methods

After IRB approval, two groups of reviewers were recruited for this study. Representing the crowd were Amazon's Mechanical Turk™ users. These users are anonymous crowd workers from diverse backgrounds who complete small tasks for remuneration on a Mechanical Turk website, and the recruiting process was completed through this website. The second group consisted of expert faculty surgery graders, recruited through email. Six hundred prequalified Mechanical Turkers™ were recruited for the study (Fig. 1). Crowd workers must have met criteria as described by Chen et al. to qualify for the study, including having previously completed 100 or more Human Intelligence Tasks (HITs), and they must have achieved a greater than 95% approval rating as qualified by the Mechanical Turk at the time of the study.[9] A HIT is simply shorthand for a single task, which is hosted on the Mechanical Turk interface. The crowd workers' identities are anonymous and users can only be identified by unique user ID codes generated by the website. Gender, age, sex, and ethnicity were not available to the authors for this study. Each crowd worker was compensated 0.75 USD for assessing an individual performance. Seven experienced robotic surgeons, each of whom rated all videos once, made up the expert group. All the surgeons are part of practices in which minimally invasive surgery is the primary technique and they all had previous experience evaluating videos of surgical performance. The surgeon group was not compensated for participating in this study.

An online survey was developed and hosted on a secure server accessible only by recruited Mechanical Turk users. The survey contained an initial qualification question in which the crowd reviewers were shown two videos, displayed side by side, of a pair of surgeons performing a Robotic Fundamentals of Laparoscopic Surgery (RFLS) block transfer task (Fig. 2). The video on the left side of the screen showed a surgeon performing the tasks with a high level of proficiency compared to the video on the right side of the

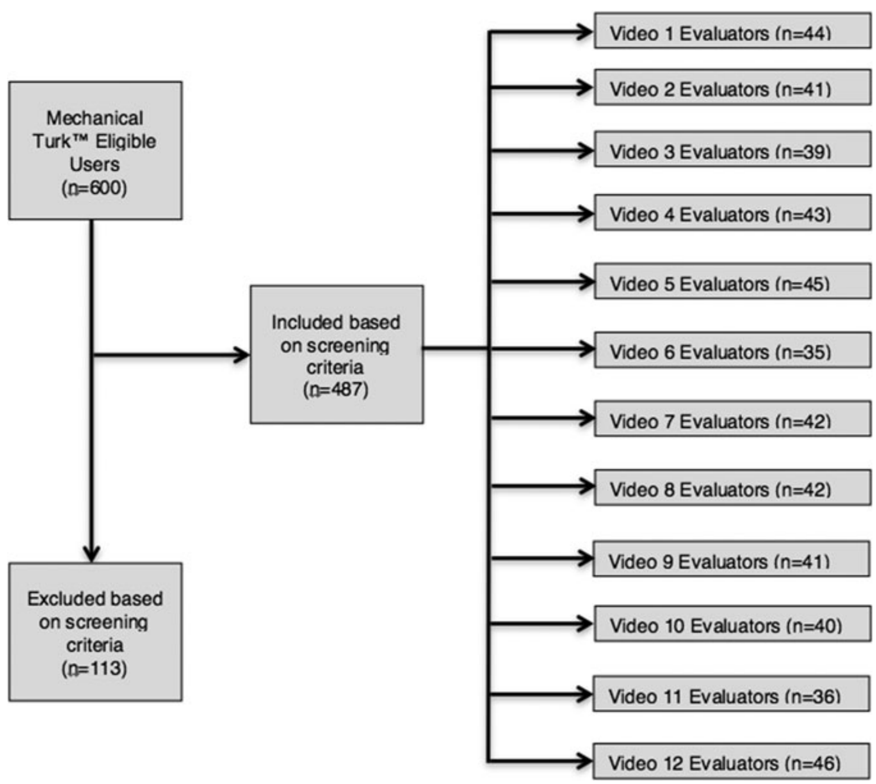


FIG. 1. Flowchart showing the breakdown of included Mechanical Turk™ graders randomly assigned to each of the 12 videos.

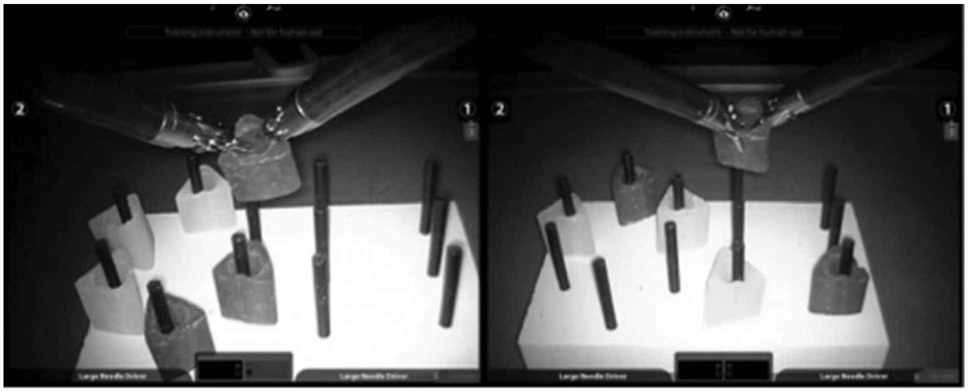


FIG. 2. Robotic fundamentals of the laparoscopic surgery (RFLS) block transfer task side by side video used to screen subjects.

screen, which showed a surgeon performing the task with an intermediate level of proficiency. These proficiency levels are based on published metric benchmarks for this particular task.[11,12] Crowd workers were asked to pick the video with the higher level of proficiency, prompting exclusion of those who answered incorrectly from the data analysis. Those excluded from the analysis were still remunerated. In addition, embedded in the survey was an attention question, which was designed so that only users who were actively paying attention to the survey would be able to correctly answer the question. Any crowd workers who answered the question incorrectly were screened out of the study and excluded from analysis (Fig. 1).

As part of the survey, we obtained recorded videos of 12 different surgeons of varying skill levels performing live porcine robot-assisted urinary bladder closures (Fig. 3). No identifying information of the surgeons performing the bladder closures was present in the videos. The length of the videos ranged between 2 and 11 minutes, with the average length being 4 hours 30 minutes. The videos were uploaded to the online survey, and evaluators were asked to evaluate the videos across five GEARS domains—bimanual dexterity, depth perception, efficiency, force sensitivity, and robotic control (Fig. 4). GEARS is an already validated tool used to assess robotic surgery.[13] Fifty unique Mechanical Turk crowd workers and seven expert surgeons evaluated each video based on the five GEARS domains. Crowd workers were only allowed to assess each performance once, but could assess more than one video if they chose to. The reason for having 50 crowd workers grade each video as opposed to larger or smaller numbers was based on a previous internal analysis of data (Chen et al.), which found 30–50 crowd responses sufficient to achieve satisfactory agreement with expert grades.[9]

Each grader's Likert ratings across each of the five GEARS domains were summed to acquire composite performance scores for each video. This yielded a composite score scale of 5–25. Means of the crowd composite scores were assessed for concordance using Cronbach's Alpha statistic (Table 1). Cronbach's Alpha scores above 0.9 indicate excellent agreement, scores from 0.9 to 0.7 indicate good agreement, and scores below 0.5 indicate poor and unacceptable levels of agreement.[14]

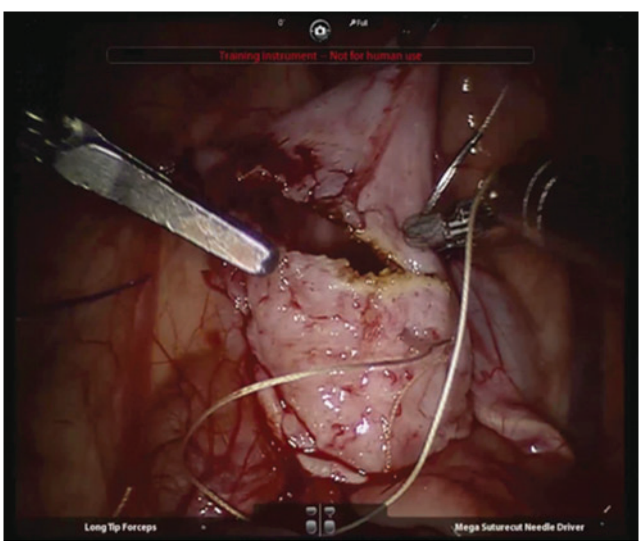


FIG. 3. Image from one of the suturing performances that was graded by both expert surgeons and Amazon's Mechanical Turk crowd.

Results

After excluding crowd workers who failed the attention or discrimination question, we were left with valid scores tfrom 487 of 600 Mechanical Turk crowd workers (Fig. 1). It took 4 hours 28 minutes to receive all crowd worker grades for the 12 videos. In comparison, it took 14 days to receive grades from all seven expert surgeons. Composite scores given by both the crowds and experts are shown in Table 1. Concordance between the surgeons and crowd was 0.93 using Cronbach's Alpha statistic, which indicates excellent agreement (Table 1). The linear relationship between the surgeon grades and crowd grades is shown in Figure 5. The R^2 value is 0.91. Standard error is shown in Figure 6.

Discussion

The current gold standard, an OSATS (Objective Structured Assessment of Technical Skills)-like method for objectively assessing surgical skill, continues to be underutilized due to cost, resource intensiveness, and the lag-time for return of results. Feedback is most effective if given immediately or near real time; therefore, existing OSATS practices tend to be deficient outside an academic research project.[15] Due to the significant variability in the absence of an agreement workshop and mentor bandwidth precluding

Depth perception

1	2	3	4	5
Constantly overshoots target, wide swings, slow to correct		Some overshooting or missing of target, but quick to correct		Accurately directs instruments in the correct plane to target

Bimanual dexterity

1	2	3	4	5
Uses only one hand, ignores nondominant hand, poor coordination		Uses both hands, but does not optimize interaction between hands		Expertly uses both hands in a complementary way to provide best exposure

Efficiency

1	2	3	4	5
Inefficient efforts; many uncertain movements; constantly changing focus or persisting without progress		Slow, but planned movements are reasonably organized		Confident, efficient and safe conduct, maintains focus on task, fluid progression

Force sensitivity

1	2	3	4	5
Rough moves, tears tissue, injures nearby structures, poor control, frequent suture breakage		Handles tissues reasonably well, minor trauma to adjacent tissue, rare suture breakage		Applies appropriate tension, negligible injury to adjacent structures, no suture breakage

Robotic control

1	2	3	4	5
Consistently does not optimize view, hand position, or repeated collisions even with guidance		View is sometimes not optimal. Occasionally needs to relocate arms. Occasional collisions and obstruction of assistant		Controls camera and hand position optimally and independently. Minimal collisions or obstruction of assistant

FIG. 4. The five Global Evaluative Assessment of Robotic Skills (GEARS) domains that were used to score the videos. Composite scores of the five domains were used to compare surgeon *vs* Turker grading.

frequent iterative trainee objective technical skills assessment, alternative methods to assist in these goals are required. In addition, video reviews may not be as objective when performed by reviewers who are within the same institution as the trainees.[16]

In Holst et al. and Chen et al., it was noted that a Crowdsourced Assessment of Technical Skills (C-SATS) was not designed to replace one on one instruction and evaluation in the setting of residency training, but may provide an adjunct method of providing quick feedback and identifying trainees who are deficient in one area of training. Traditional methods of instruction and feedback are invaluable because they offer content expertise and transfer information about the nuances of surgery that could not be yielded by crowds; however, C-SATS may have a role in rapidly triaging trainees with deficiencies and allowing mentors to target valuable training resources to these deficiencies, as opposed to teaching all trainees with the same curricula. Feedback from crowds may be obtained rapidly enough to provide this guidance between surgical cases or between days in the operating room.

C-SATS has been used in a residency training environment, which is ideally suited to this method because of the

Table 1. Summary of Grades Assigned to Each of the 12 Video Performances

	Mechanical Turk™ graders				*Surgeon graders*		
	Initial, N	*Qualified,* N	*C-SATS mean (SD)*	*95% CI*	*Number of graders, N*	*C-SATS mean (SD)*	*95% CI*
Video 1	50	37	21.49 (3.42)	± 1.10	7	18.71 (1.67)	± 2.99
Video 2	50	41	20.95 (3.81)	± 1.17	7	18.00 (3.39)	± 2.96
Video 3	50	39	20.36 (3.51)	± 1.10	7	16.57 (5.39)	± 3.57
Video 4	50	43	18.02 (4.69)	± 1.40	7	15.85 (3.21)	± 3.01
Video 5	50	45	20.29 (3.28)	± 0.96	7	17.85 (5.10)	± 3.91
Video 6	50	35	20.37 (3.56)	± 1.18	7	18.14 (3.85)	± 2.58
Video 7	50	42	20.02 (4.04)	± 1.22	7	16.29 (5.72)	± 3.82
Video 8	50	42	21.45 (2.74)	± 0.83	7	17.85 (3.59)	± 2.07
Video 9	50	41	15.10 (4.87)	± 1.49	7	10.71 (1.92)	± 2.52
Video 10	50	40	17.13 (3.78	± 1.17	7	11.57 (2.05)	± 2.49
Video 11	50	36	18.47 (4.84)	± 1.58	7	14.57 (2.88)	± 1.76
Video 12	50	46	15.48 (4.43)	± 1.28	7	9.00 (1.67)	± 1.47
Cronbach's Alpha		0.93					

C-SATS = Crowd-sourced Assessment of Technical Skills; SD = standard deviation; CI = confidence interval.

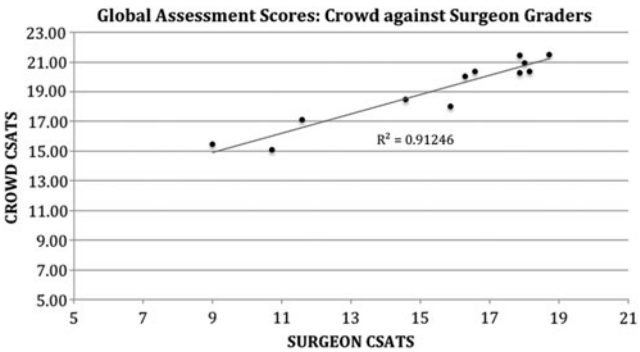


FIG. 5. Crowd-sourced Assessment of Technical Skills (C-SATS): Global performance Scores provided by the crowd against the global performance scores provided by the expert surgeon graders. The R^2 value of the best fit line is 0.91.

controlled learner-centered nature of residency. Holst et al. showed that crowds can identify differences in urology resident training levels and that crowdsourcing is a practical effective way of providing feedback in near real time.[10] The major limitation of that study, however, was that all tasks evaluated were dry-laboratory tasks. In a setting of resident work-hour restrictions, surgical trainees are spending more time in simulation laboratories to refine their technical skills, and thus, it is important that crowds can evaluate these dry-laboratory tasks quickly; however, it is vital to prove that crowds can also judge technical skill being performed on live tissue as opposed to dry-laboratory materials. Animate surgery better approximates real human surgery; thus, our hypothesis needed to be tested in this environment as a next step in validating C-SATS. With no knowledge of relevant anatomy, crowds provided extremely rapid and accurate feedback in comparison to expert graders.

A limitation of this study is that only one type of live-tissue performance was assessed and the surgery was still in a controlled environment through a porcine laboratory. In addition, all videos assessed were relatively short (averaging under 5 minutes in length). It remains to be seen if crowd evaluators can continue to provide effective grading across a range of live-tissue surgeries with varying lengths. Future studies aim to include videos across a range of surgical approaches, such as laparoscopic and open surgeries. While additional validation is needed before C-SATS is embedded into training centers, evidence that crowds can evaluate live-tissue surgery adds to the growing body of literature for the value of this adjunctive objective assessment tool.

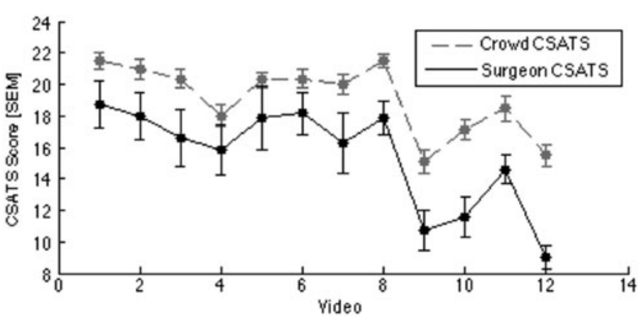


FIG. 6. The mean score of each video (*circle*) is provided for the crowd and surgeon C-SATS groups along with error bars for the standard error of the mean to indicate variation of the mean within our data.

Another limitation to this study is that the performances assessed were from a wide range of surgical skill levels from robotic faculty to novice trainees. Thus, the skill effect-size may have been disparate enough for lay people to easily see differences. It is arguable that if the cohort of performers were of more similar skill levels, it would require expert observers to discriminate the smaller technical skills differences. Resident training environments where the skills of the trainees vary significantly are ideally suited for using this methodology. Additional studies will be needed to test C-SATS on cohorts of surgeons who have similar skills.

Conclusion

We demonstrate that crowdsourcing basic surgical skills of animate surgery compares favorably to a panel of expert surgeon assessors and is faster than the experts—providing large-volume feedback in a matter of hours. Utilizing crowdsourcing as a means to assess technical surgical skills provides an inexpensive, scalable, rapid, and effective way to evaluate live-tissue procedures, paving the way for further validation in human surgery. Ultimately, C-SATS assessments will need to be linked to clinical outcomes to gain confidence that presumably nonmedically trained crowds of people can accurately ascribe surgical skill.

Author Disclosure Statement

Drs. L.W.W., T.M.K., and T.S.L. are now equity shareholders in the company CSATS, Inc., which is a company spun out of the University of Washington's Technology Transfer Office to commercialize CSATS. However, all material and data presented in this article took place before the formation of CSATS, Inc., and thus represent efforts made without the umbrella of financial incentive.

References

1. Zhan C, Miller MR. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. JAMA 2003;290:1868–1874.
2. Rogers SO, Jr., Gawande AA, Kwaan M, et al. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. Surgery 2006;140:25–33.
3. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med 2013;369:1434–1442.
4. Van Hove PD, Tuijthof GJM, Verdaasdonk EGG, et al. Objective assessment of technical surgical skills. Br J Surg 2010;97:972–987.
5. Datta V, Bann S, Mandalia M, et al. The surgical efficiency score: A feasible, reliable, and valid method of skills assessment. Am J Surg 2006;192:372–378.
6. Anon. Crowdsourcing. Merriam-Webster.com. Merriam-Webster, n.d. Web. Available at: www.merriam-webster.com/dictionary/crowdsourcing. Accessed January 17, 2015.
7. Bigham JP, Jayant C, Ji H, et al. VizWiz: Nearly real-time answers to visual questions. In: Proceedings of the 23rd

Annual ACM Symposium on User Interface Software and Technology. UIST'10. New York, NY: ACM. 2010 pp 333–342. Available at: http://doi.acm.org/10.1145/1866029.1866080. Accessed May 3, 2014.
8. Khatib F, Cooper S, Tyka MD, et al. Algorithm discovery by protein folding game players. Proc Natl Acad Sci 2011. Available at: http://www.pnas.org/content/early/2011/11/02/1115898108. Accessed May 3, 2014.
9. Chen C, White L, Kowalewski T, et al. Crowd-Sourced Assessment of Technical Skills: A novel method to evaluate surgical performance. J Surg Res 2014;187:65–71.
10. Holst D, Kowalewski TM, White LW, et al. Crowd-sourced assessment of technical skills: An adjunct to urology resident surgical simulation training. J Endourol 2014 [Epub ahead of print].
11. Tausch TJ, Kowalewski TM, White LW, et al. Content and construct validation of a robotic surgery curriculum using an electromagnetic instrument tracker. J Urol 2012;188: 919–923.
12. Lendvay TS, Hannaford B, Satava RM. Future of robotic surgery. Cancer J 2013;19:109–119.
13. Goh AC, Goldfarb DW, Sander JC, et al. Global evaluative assessment of robotic skills: Validation of a clinical assessment tool to measure robotic surgical skills. J Urol 2012;187:247–252.
14. Cronbach LJ, Shavelson RJ. My current thoughts on coefficient alpha and successor procedures. Educ Psychol Meas 2004;64:391–418.
15. Moorthy K, Munz Y, Sarker SK, et al. Objective assessment of technical skills in surgery. BMJ 2003;327:1032–1037.
16. Darzi A, Smith S, Taffinder N. Assessing operative skill. Needs to become more objective. BMJ 1999;318:887–888.

Address correspondence to:
Daniel Holst, BS
University of Washington School of Medicine
3824 Fremont Lane N.
Seattle, WA 98103

E-mail: dholst@uw.edu

Abbreviations Used

C-SATS = Crowd-sourced Assessment of Technical Skills
GEARS = Global Evaluative Assessment of Robotic Skills
GOALS = Global Operative Assessment of Laparoscopic Skills
OSATS = Objective Structured Assessment of Technical Skills
HIT = Human Intelligence Task
RFLS = Robotic Fundamentals of Laparoscopic Surgery

J Robotic Surg (2016) 10:49–56
DOI 10.1007/s11701-015-0540-1

ORIGINAL ARTICLE

Urology residents experience comparable workload profiles when performing live porcine nephrectomies and robotic surgery virtual reality training modules

Vladimir Mouraviev[1] · Martina Klein[2] · Eric Schommer[3] · David D. Thiel[3] · Srinivas Samavedi[1] · Anup Kumar[1] · Raymond J. Leveillee[4] · Raju Thomas[5] · Julio M. Pow-Sang[6] · Li-Ming Su[7] · Engy Mui[8] · Roger Smith[9] · Vipul Patel[1]

Received: 13 October 2015 / Accepted: 15 October 2015 / Published online: 11 January 2016

Abstract In pursuit of improving the quality of residents' education, the Southeastern Section of the American Urological Association (SES AUA) hosts an annual robotic training course for its residents. The workshop involves performing a robotic live porcine nephrectomy as well as virtual reality robotic training modules. The aim of this study was to evaluate workload levels of urology residents when performing a live porcine nephrectomy and the virtual reality robotic surgery training modules employed during this workshop. Twenty-one residents from 14 SES AUA programs participated in 2015. On the first-day residents were taught with didactic lectures by faculty. On the second day, trainees were divided into two groups. Half were asked to perform training modules of the Mimic da Vinci-Trainer (MdVT, Mimic Technologies, Inc., Seattle, WA, USA) for 4 h, while the other half performed nephrectomy procedures on a live porcine model using the da Vinci Si robot (Intuitive Surgical Inc., Sunnyvale, CA, USA). After the first 4 h the groups changed places for another 4-h session. All trainees were asked to complete the NASA-TLX 1-page questionnaire following both the MdVT simulation and live animal model sessions. A significant interface and TLX interaction was observed. The interface by TLX interaction was further analyzed to determine whether the scores of each of the six TLX scales varied across the two interfaces. The means of the TLX scores observed at the two interfaces were similar. The only significant difference was observed for frustration, which was significantly higher at the simulation than the animal model, $t\ (20) = 4.12$, $p = 0.001$. This could be due to trainees' familiarity with live anatomical structures over skill set simulations which remain a real challenge to novice surgeons. Another reason might be that the simulator provides performance metrics for specific performance traits as well as composite scores for entire exercises. Novice trainees experienced substantial mental workload while performing tasks on both the simulator and the live animal model during the robotics course. The NASA-TLX profiles demonstrated that the live animal model and the MdVT were similar in difficulty, as indicated by their comparable workload profiles.

Keywords Robotic surgery training · Mental workload · NASA Task Load Index

Introduction

Robotic-assisted urologic surgery is predicted to continue to grow in usage in the coming years, and residents trained in urology will increasingly be expected to be proficient in

✉ Vladimir Mouraviev
Vladimir.Mouraviev@flhosp.org

1 Global Robotics Institute, Celebration, FL, USA

2 Department of Psychological Sciences, Texas Tech University, Lubbock, TX, USA

3 Department of Urology, Mayo Clinic, Jacksonville, FL, USA

4 Department of Urology, University of Miami, Coral Gables, FL, USA

5 Department of Urology, Tulane University Medical Center, Tulane, LA, USA

6 Department of Urological Oncology, Moffitt Cancer Center, Tampa, FL, USA

7 Department of Urology, University of Florida, Gainesville, FL, USA

8 University of Central Florida College of Medicine, Orlando, FL, USA

9 Nicholson Center, Florida Hospital, Celebration, FL, USA

robotic surgery [1]. The complexity of robotic technology, its steep learning curve, and work-hour limitation of resident trainees make incorporating robotic training into residency a challenging task. Experts suggest that learning as a bedside assistant for robotic surgery has a rapid plateau; many programs are now utilizing physician assistants and surgical technicians for bedside duties to free the residents for console training [2]. In high volume programs it remains difficult for residents to gain hands-on console time due to their insufficient skill set and the complexity of most procedures.

Robotic simulation training tools can, therefore, be utilized by novice trainees to shorten the learning curve and improve operative skills in a low-risk environment. In pursuit of improving the quality of residents' education, the Southeastern Section of the American Urological Association (SES AUA) hosts an annual robotic training course for its residents. This workshop involves training of basic laparoscopic surgery skills using virtual reality training modules of the Mimic da Vinci-Trainer (MdVT, Mimic Technologies, Inc., Seattle WA, USA) as well as training on performing a nephrectomy using a live porcine model. For simulation training to be successful, it is essential that it (1) practices the relevant skills and (2) matches the level of difficulty (workload demand) similar to the demands experienced in the "real" procedure. Thus, the goal for the present study was to assess whether the workload demands experienced in the virtual simulation training environment, which trains basic robotic surgery skills, match those experienced in when performing the live nephrectomy using a porcine model.

Materials and methods

Select residents from each of the 14 training programs of SES AUA were invited to Orlando, FL, for a 2-day robotics training course. Up to 3 residents were invited from each training program, and 21 participated in the training course. This cohort of residents represented a wider range of training and diversity in experience than in previous courses, being exposed to robotic surgery early at their home institutions. Volunteer faculty were recruited from SES AUA training programs.

The 2015 annual SES AUA robotics training workshop, which is outlined in more detail below, involved training nephrectomy on a porcine model as well as training on the MdVT trainer. Participants' workload we assessed at both interfaces (MdVT and live porcine model) using the NASA Task Load Index (NASA-TLX). The NASA-TLX assesses workload along six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration [3]. Each is measured on a 21-point scale; scores can range between 0 ("Very Low") and 100 ("Very High"), see "Appendix 2".

The SES AUA robotics course is outlined below [4].

Robotic course day 1

A full didactic session was broken into three components. Component 1 covered the basics of robotic surgery including room set-up, bedside assistance, and console essentials. Component 2 covered several aspects of robotic kidney surgery including patient positioning, port placement, and surgical techniques. Component 3 focused on robotic prostate surgery including port placement and different surgical techniques. Didactics were supplemented with surgical videos and discussions of difficult surgical scenarios and possible complications.

Robotic course day 2

The trainees were divided into two groups. Half were asked to perform skill tasks on the Mimic da Vinci-Trainer (MdVT, Mimic Technologies, Inc., Seattle, WA, U.S.A) for 4 h using the dV trainer (version 2) while the other half performed set tasks in a live nephrectomy on porcine model using the da Vinci Si robot (Intuitive Surgical Inc., Sunnyvale, CA). After 4 h the groups changed places for another 4-h session.

Simulation section

In the 4-h MdVT simulation session, trainees were first given a tutorial of the console and its functionality. The trainees then proceeded to complete five exercises with increasing difficulty and required skills. The first exercise, "pick and place", involved simple movements of rings from one pole to another and is used to orient the trainee to the simulator. The second exercise, "peg board" is more advanced and required the trainee to clutch hand instruments while moving the camera, which involves coordinated hand and foot movements. The third exercise, "ring walk", involved moving a ring over a curvy bar without touching the bar with any portion of the ring. This drill requires all the above skills as well as maintaining awareness and accuracy with the ring position at all time. The fourth exercise, "thread the rings", involves passing a curved needle through rings positioned at different angles without touching the ring with any part of the needle. This drill teaches trainees good suturing technique. The last exercise, "tubes 2", is the most challenging and realistic. This drill is designed to replicate performing an urethrovesical anastomosis. It utilizes all of the above skills including accuracy, coordination, and sufficient needle control.

Animal training section

In the 4-h porcine model live surgery session, all trainees spent 1 h performing cystostomies and cystorrhaphies. They then spent 30 min practicing port insertion and robot docking. Finally, for 2.5 h, trainees conducted a bilateral nephrectomy which included artery, vein, kidney and ureter dissections and ligation.

Questionnaire

All trainees were asked to complete the NASA-TLX 1-page questionnaire following both the MdVT simulation and live animal model sessions.

Results

Twenty-one residents from 14 programs in the SES AUA participated in this course. Seventeen (80.9 %) had used a console during an actual surgical case, while four did not. The distribution of the different levels of training among the residents is shown in Fig. 1. Unlike previous years' courses when only senior or chief residents participated, this course included more junior residents. This reflects a shift toward early exposure to robotic surgery during urology training in most academic programs. The number of robotic surgeries performed or assisted by residents at different levels of training is shown in Fig. 2. Trainees' satisfaction with their program robotic surgery training was assessed (Fig. 3). Of the 17 residents who performed actual robotic surgery, 7 (41.2 %) stated that the simulator replicates real-life robotic surgery, while 10 (58.8 %) stated that it did not.

The NASA-TLX scores were converted to a 0–100 scale with 5-point increments. The raw TLX method was employed to eliminate the weighting process of the different TLX scales. To assess the NASA-TLX data at two interfaces (simulator vs. animal model) for the different levels of training (year of residency), a 4 (training level) × 2 (interface) × 6 (TLX scales) mixed ANOVA was computed. The Greenhouse-Geisser correction was used to correct for violations of the sphericity assumption. The ANOVA indicated a significant main effect for TLX scales, F (3.91, 66.44) $= 4.93$, $p = 0.002$, $\eta_{\text{partial}} = 0.225$, as well as a significant interface by TLX scales interaction, F (3.73, 63.42) $= 3.73$, $p = 0.016$, $\eta_{\text{partial}} = 0.166$). None of the other main effects and interactions were significant. To further analyze the TLX main effects, Bonferroni-corrected repeated-measures t tests were computed to determine which TLX scales differed significantly from each other; type-I error rate per comparison was set to 0.003. Means of the TLX scales are presented in Fig. 4. As can be seen from Fig. 4, effort resulted in the highest score. The Bonferroni-corrected t-tests indicated that mental demand was significantly higher than physical demand [t (20) $= 4.05$, $p = 0.001$] and then frustration

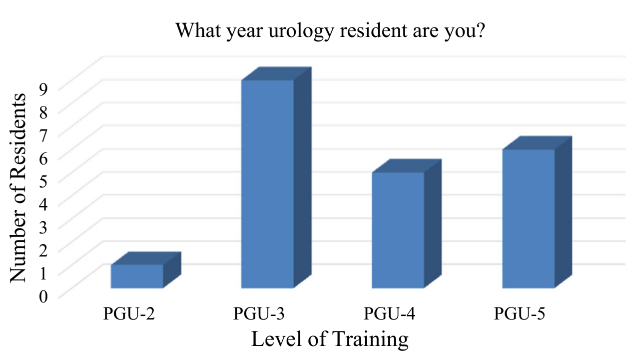


Fig. 1 Robotic Simulator Questionnaire: question 1 results

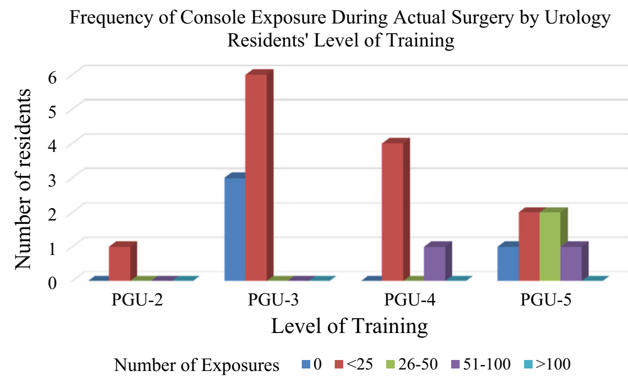


Fig. 2 Robotic Simulator Questionnaire: question 4 results

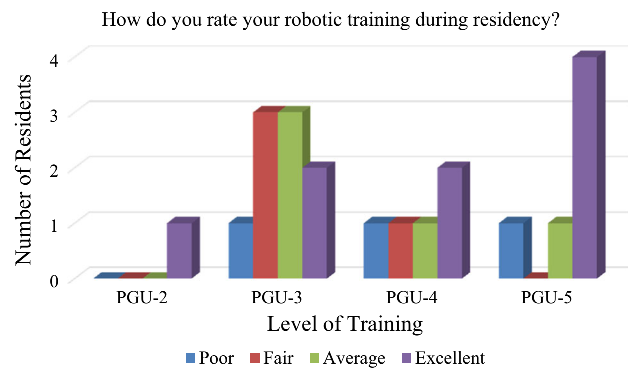


Fig. 3 Robotic Simulator Questionnaire: question 5 results

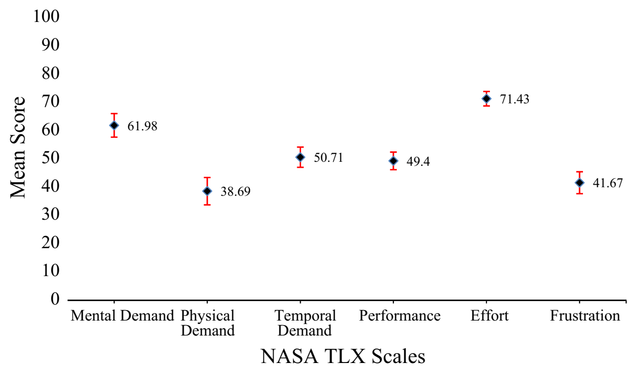


Fig. 4 Mean scores of the NASA-TLX scales. *Error bars* refer to standard error of the mean

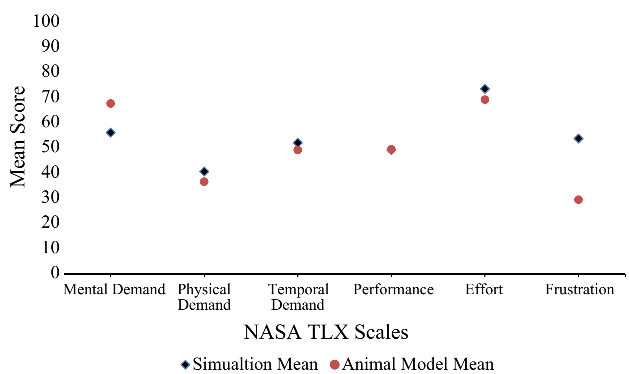


Fig. 5 Mean scores of the NASA-TLX scales in simulation versus animal model

[t (20) = 3.52, p = 0.002]. Further, temporal demand was significantly higher than physical demand [t (20) = 2.90, p = 0.009] and effort was significantly higher than physical demand [t (20) = 6.52, $p < 0.001$)], temporal demand [t (20) = 5.12, $p < 0.001$], performance [t (20) = 5.15, $p < 0.001$], and frustration [t (20) = 6.90, $p < 0.001$].

The analysis of the interface by TLX interaction was further analyzed to determine whether the scores of each of the six TLX scales varied across the two interfaces. On that end, Bonferroni-corrected repeated-measures t tests were computed; type-I error rate per comparison was set at $\alpha = 0.008$. The means of the TLX scores observed at the two interfaces are in Fig. 5. The only significance was observed for frustration, which was significantly higher at the simulation than the animal model, t (20) = 4.12, p = 0.001.

Discussion

Robotic surgery is increasing in popularity in the field of urology due to its minimal invasiveness, reduced risk of complications, and shortened hospital stay. This growing trend is evident in our results. The majority of the trainees this year (80 %) reported live console exposure. In contrast with a similar survey conducted in 2013 in a group of SES AUA trainees, only 56.9 % of the trainees that year reported having had robotic console time [4]. During the 2014 annual training course 92 % of the trainees reported performing live robotic surgery at their home institution [5]. Despite these increasing numbers, there is a lack of standardization and certification process for urology residents in robotic surgery. Furthermore, there is no standardized training protocol for residents learning robotic surgery across the various training programs. Gover et al. suggested a threshold of 25–30 cases for a novice surgeon to begin to operate the foot pedals and controls safely and intuitively [6]. Only 4 (19 %) of our trainees reported having performed more than 25 cases.

Robotic surgery simulators have been proposed to narrow the gap of novice trainees' skill levels [7]. Such simulators can help establish the basics of important operative skills such as eye–hand–foot coordination and using the console controls and foot pedals. Our program chose to use the MdVT simulator for training. The Mimic da Vinci-Trainer (MdVT, Mimic Technologies, Inc., Seattle, WA, and USA) is one of the most established virtual robotic surgical simulators today. Previous research indicated that training on the MdVT resulted in superior surgical performance compared to solely training on the real da Vinci surgical system (Intuitive Surgical Inc., Sunnyvale, CA) when taking a robotic skills assessment using the real da Vinci system [9].

The goal for the present study was to determine whether performance of the robotic surgery simulator tasks employed by the training course of the SES AUA matches the workload demands when performing a real robotic surgery. Towards that end, a porcine nephrectomy was employed. Thus, the results of the present study indicate that the simulation exercises employed by SES AUA generally induce similar workload demands to those experienced when performing a live porcine nephrectomy, indicating that the simulation exercises are not too easy. Specifically, the results indicated that mental demand and effort were major contributors of workload across both surgical interfaces. Further, the

different workload dimensions did not significantly differ across the two surgical interfaces, with the exception of frustration. Significantly higher frustration levels were observed at the simulation than the animal. Higher frustration might be due to trainees being more familiar with alive anatomical structures than the simulation exercises. Another potential reason for the simulation to induce higher frustration levels than the animal is that the simulator provides metrics for specific performance traits, as well as a composite performance score [7]. In addition to the objective metrics, the MdVT simulator defines thresholds which indicate whether the trainee's score is considered a "passing" or "failing" performance with acceptable and warning scoring levels, respectively [7]. Conversely, the animal hands-on part did not have objective metric parameters to assess the skill set of trainees in robotic surgery. The faculty of the course subjectively evaluated the proficiency levels of residents when they performed the porcine nephrectomy. Furthermore, the timeframe for every trainee was limited at the robotic console when performing the nephrectomy when compared to the simulation.

However, though training on the MdVT simulator, has been validated [8], its use is not without limitations. There is an initial purchasing cost which ranges from $85,000 to $100,000. These are added costs of annual maintenance fees. There are currently no urology specific procedure modules or simulation drills available but only general surgical skill tasks like the ones used during the SES AUA training course. This limitation could hinder a rapid learning plateau and might not translate to better operative skills without supplementing with real live surgery console time. Therefore, work on more realistic 3D case simulations to advance clinical decision-making and procedural knowledge is currently in progress. The animal lab used for the course in this analysis cost roughly $1,900/h for the animal models, pharmaceuticals, veterinary support, robotic equipment with instruments, PPE, and the specially equipped facility. Other sites have reported $500/h, but this only includes the cost of the animal model, not the entire package of services and equipment [10]. It also lacks realistic human anatomy and might provide a false sense of security which could lead to harming a patient [11]. Future work should be invested in developing urology-specific training modules such as radical prostatectomy and partial nephrectomy simulations. The existing application only hones skills used in general robotic surgery and is not necessarily reflective of skills needed to perform urologic robotic surgery.

Educators and companies have yet to determine the best model to use for teaching robotic surgery. Many factors must be taken into consideration including the cost, availability of expert faculty, legal responsibility on such supervising faculty, risk to patients, and the additional workload on trainees.

These results of the present study, combined with previous and future SES AUA training course results, can significantly enhance our efforts to establish a standardized robotic surgery training program that is cost-effective, practical, and of the highest quality. Encouraging the development of urology-specific robotic training tools in simulation will also aid in reaching our goal. Some limitations of this analysis include its regional focus and limited sample size. It surveyed a limited number of trainees from the SES AUA and is not representative of trainees across the country. The analysis also did not assess the methods each program uses for robotic training. Upon completion of the residency program, many urologists recognize the effort and learing curve involved in acquiring robotic surgery skills and arrive at a consensus that training and proficiency in robotic surgery are necessary during residency [9]. Future direction for this project includes compiling detailed accounts of trainees' exposures at their home institutes. Such analysis combined with future performance scores and trainees' subjective opinions could lead to identifying the most effective methods of training. Work is currently in progress to improve the current robotic training methods.

Conclusions

Trainees experienced similar levels of workload when performing the virtual reality training modules and when performing a live porcine nephrectomy, indicating that the MdVT virtual reality training modules employed by SES AUA workshop have adequate difficulty.

Acknowledgments The authors would like to cordially thank members of Global Robotic Institute Mrs. Ashley Fialkowski and Mrs. Kim Straw for their assistance in obtaining the questionnaires from participants.

Compliance with ethical standards

Conflict of interest Dr Vipul Patel is a consultant for MIMIC Technologies, Inc.

Appendix 1: Robotic Simulator Questionnaire

1. What year urology resident are you?
   - ☐ Uro-1
   - ☐ Uro-2
   - ☐ Uro-3
   - ☐ Uro-4
   - ☐ Uro-5
2. Does your training program own or have access to a robotics simulator?
   - ☐ No
   - ☐ Mimic Simulator
   - ☐ Ross Simulator
   - ☐ Mimic Backpack or console
   - ☐ Other__________
3. Have you been on the robotics console for an actual case?
   - ☐ Yes
   - ☐ No
4. Approximate the number of cases on which you have robotics console time
   - ☐ <25
   - ☐ 26-50
   - ☐ 51-100
   - ☐ >100
5. How do you rate your robotic training during residency?
   - ☐ Poor
   - ☐ Fair
   - ☐ Average
   - ☐ Excellent
6. In your experience, do you feel that the simulator replicates real life robotics?
   - ☐ Yes
   - ☐ No
7. Which drill did you find the most difficult?
   - ☐ Peg board
   - ☐ Ring Walk
   - ☐ Thread the rings
   - ☐ Tubes 2
8. If your program lacks a robotics simulator, do you think this device would be helpful in your program?
   - ☐ Yes
   - ☐ No

Appendix 2: NASA TASK Load Index

Name	Task	Date

Mental Demand How mentally demanding was the task?

Very Low Very High

Physical Demand How physically demanding was the task?

Very Low Very High

Temporal Demand How hurried or rushed was the pace of the task?

Very Low Very High

Performance How successful were you in accomplishing what you were asked to do?

Perfect Failure

Effort How hard did you have to work to accomplish your level of performance?

Very Low Very High

Frustration How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low Very High

References

1. Duchene DA, Rosso F, Clayman R, McDougall EM, Winfield HN (2011) Current minimally invasive practice patterns among postgraduate urologists. J Endourol Endourol Soc 25(11):1797–1804
2. Thiel DD, Lannen A, Richie E, Dove J, Gajarawala NM, Igel TC (2013) Simulation-based training for bedside assistants can benefit experienced robotic prostatectomy teams. J Endourol Endourol Soc 27(2):230–237
3. Hart SGNASA (2006) Task Load Index (NASA-TLX); 20 years later. Proc Human Factors Ergonom 50:904–908
4. Thiel DD, Patel VR, Larson T, Lannen A, Leveillee RJ (2013) Assessment of robotic simulation by trainees in residency programs of the southeastern section of the American urologic association. J Surg Educ 70(5):571–577
5. Ruparel RK, Taylor AS, Patel J et al (2014) Assessment of virtual reality robotic simulation performance by urology resident trainees. J Surg Educ 71(3):302–308
6. Grover S, Tan GY, Srivastava A, Leung RA, Tewari AK (2010) Residency training program paradigms for teaching robotic surgical skills to urology residents. Curr Urol Rep 11(2):87–92
7. Smith R, Truong M (2013) Simulation in robotic surgery: a comparative review of simulators of the da Vinci. Modelbenders Press, Oviedo, FL
8. Lerner MA, Ayalew M, Peine WJ, Sundaram CP (2010) Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system? J Endourol Endourol Soc 24(3):467–472
9. Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A (2009) Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. Urology 73(6):1288–1292
10. Perrenot C, Perez M, Tran N et al (2012) The virtual reality simulator dV-Trainer((R)) is a valid assessment tool for robotic surgical skills. Surg Endosc 26(9):2587–2593
11. Hoffman MS (2012) Simulation of robotic hysterectomy utilizing the porcine model. Am J Obstet Gynecol 206(6):523 e521–522

Response to "Unlike History, Should a Simulator Not Repeat Itself?" Simulation in Healthcare 2015; 10(6): 331–335

To the Editor:

I have read with great interest Dr Lampotang's editorial in the December 2015 issue of *Simulation in Healthcare.*[1] The author has provided an excellent overview of some of the benefits of and the difficulty in achieving repeatability in health care training simulations. The categories and examples included should become common references for our unique niche of the simulation community in the years to come. The topic of repeatability has also been actively investigated in interactive, networked, and parallel simulation systems, often associated with military training applications. That community has literally created hundreds of simulation systems to address various problems and found repeatability to be important in applications like analytic war games, which need to be run hundreds of times with very controlled differences in actions but without uncontrolled variations from internal algorithms or data transfer times. Distributed simulation events that link multiple simulators via computer networks also encounter undesired repeatability issues primarily from two causes, differences in message delivery times from one run to another and the internal logic used to sequence events received from external systems, which all have the same logical simulation time. These simulation communities have developed algorithms that specifically control for these variations and software infrastructures that attempt to provide these capabilities as a service to any simulator that uses them.[2,3]

The definition that the author provides for repeatability is concise and useful from the perspective of the human users of the simulation. "Repeatability is the measure of the similarity in the outputs of a simulator during repeated runs of a given scenario with identical inputs, interventions, and events at the exact same times." This correctly identifies the fact that identical output relies on identical input in all of its forms. A simulation system of any significant complexity is prone to a large number of uncontrolled input factors and internal operations, which make repeatability extremely difficult to achieve. This response will present some of the most common of these.

When a simulation is a closed, single computer system that is driven only by preloaded digital data files, achieving repeatability is relatively straightforward. These systems can be said to be deterministic and can be structured to provide perfect repeatability, just as a calculator provides perfectly repeatable answers to the same problem every time. However, when a simulation is part of an interactive, real-time experience that includes input from external systems such as human participants and other computer devices, repeatability is much more difficult to insure and may be impossible.

For complex systems such as these, repeatability can be explored at multiple levels. Lampotang explicitly identifies the model, simulator, and simulated environment.[1] He also provides examples of the information delivery between two devices, but does not list it with the other three. He has given several excellent examples where the linking of multiple devices and the interfaces between them can create uncontrolled variation which leads to non-repeatable outcomes. These various sources might more clearly be identified as stemming from *external systems* such as the humans, computers, and devices that are part of the simulated environment; *information delivery* which includes computer networks or physical delivery lines that carry data or physical triggers to a simulator; *internal interpretation* by logic within the simulator that is used for managing and scheduling events that are internally generated or received from external systems; and *internal models* that use algorithms which may or may not provide repeatable results. Describing these sources or variation has been attempted in previous publications, though without the explicit terminology provided here.[2–4] Examples of how each of these can impact repeatability are provided below.

When seeking repeatability, it is necessary to understand the basis of the internal models or algorithms which perform the computations within the simulator. In many fields, the lack of perfect knowledge of the domain (e.g. the human body) has led to the use of stochastic and statistical models to represent the richness and diversity of the domain. These models usually rely on a random number generator (RNG) as a source of input data. As the author points out, various races respond differently to anesthesia, as do different sexes, and body masses. The details needed to model this deterministically are often unavailable or too complex to include in a simulator. In these cases, simplified tables of average responses and standard deviations around those averages are often used along with stochastic and statistical algorithms which create variability within these defined limits.[5] Together these create a simulator in which a 40 year-old, Caucasian male, 6'0", with a BMI of 25 does not always respond exactly the same to a volume of anesthetic, but always responds within known ranges for a person of that type. Such variability may be desirable for realism and uniqueness of training events, but is undesirable when repeatability is a goal.

RNGs come in many forms, some provided as software libraries and others cleverly contrived by software programmers. In all cases, these actually generate pseudo-random numbers with a demonstrable level of bias or skew. Avoiding all use of RNGs and algorithms that depend on them is one step toward creating a repeatable simulation at the model level. RNGs found in software libraries often make use of a "seed number" which kicks off a long sequence of random numbers throughout the execution of the simulation. In these cases, deliberately using the same seed number at the beginning of every simulation event will lead to the same sequence of pseudo-random numbers throughout the event. However, this apparent repeatability can be thwarted by human actions and by system behaviors during a run, as will be explained.

Inputs from an external system can also result in different computational

DOI: 10.1097/SIH.0000000000000169

outcomes. These systems may be both human users and external, networked computer devices. In both cases, the events which are generated externally and become inputs to a simulation can contain varying content which will throw off the repeatability of the simulator. These variations are much more common and extreme when the external system is a human trainee or instructor, but also occur when they originate from another simulator or device. When the events contain different contents due to slightly different actions taken externally, this information can easily lead to different decisions in the receiving simulation, change its internal state variables, and pass that variation on to the human trainee in its output. For example, when an injection of adrenalin is required to stimulate the heart, if the injection is provided only one minute later from one trial to the next, the simulator may cross an internally programmed threshold in the software. When the adrenalin is applied before this threshold is reached the patient may live, when applied afterward the patient may die, even when the difference in the 2 times is only a few seconds. Such an extreme boundary may not be typical of most real-life situations, but it is very common for software algorithms and data tables, which are programmed into all types of simulations, creating these types of hard thresholds.

Variation can also be triggered when external events are received with exactly the same internal content but arriving at a different time or in a different order. When this occurs, the simulator may receive and process event B before event A ($B < A$), rather than $A < B$ as in a previous run. This reversal of order can be caused for multiple reasons, most commonly not only because the events were actually generated at a different time by a human user, but also when a different computer system is not strictly synchronized and can send events at different times during a second or a third run of the simulation scenario. Moreover, when two or more simulators are linked together electronically, if one simulator generates multiple events at the same simulated time, these events may be delivered to another simulator in the same order, but because they have the same time stamp on them, they can logically and correctly be processed in either sequence $A < B$ or $B < A$, which contributes to nonrepeatability. If a controlled RNG is being used as described earlier, even the use of the same seed number on multiple runs cannot prevent this reversal of event order from reversing the application of the RNGs, which were used on a previous run. There are several advanced parallel and distributed simulation infrastructures that can be used to insure that multiple simulators are synchronized and events are always processed in the same order.[3,6] These include infrastructure software such as SPEEDES, which was developed by the NASA Jet Propulsion Laboratories and is available as a product from WarpIV Technologies, and the High Level Architecture (HLA) Runtime Infrastructure (RTI), which was designed by the US Department of Defense and is available as a product from multiple companies (eg, VT MAK Inc, Pitch Technologies Inc). These can eliminate variation within computer simulators, but they cannot correct or control the variation caused by human input.

Medical simulation often looks to the military as a front-runner in simulation techniques and technologies. Flight simulators, which were cited by the author and are widely understood by the public, are actually some of the most rudimentary and straightforward of these systems. As the author points out, these simulate the behavior of machines that have been engineered by humans and are understood much better than the human body. However, there are many military simulators that include models of human behavior (eg, OneSAF and SOAR), acting as individuals and groups, all of which wrestle with complexities similar to that found in modeling human physiology. Attempts to represent these behaviors have led directly to the creation of new fields of study or have expanded on existing fields—such as stochastic modeling, agent-based modeling, artificial intelligence, and machine learning. Some of these techniques may be useful in modeling human physiology and its response to various external medical and trauma stimuli.

In summary, when humans and other sources of external stimuli are part of a simulation driven event, there are almost always sources of variation that can and will lead to nonrepeatable runs of the scenario. Controlling as much of the externally generated stimuli as possible is the best option for approaching repeatability. It can lead to scenarios that are indistinguishable from each other most of the time, even though their internal state variables may have many differences. However, on occasion, these differences will cross important thresholds, which will lead to very different outcomes. True repeatability requires a level of control of the internal models, internal interpretation, information delivery, and external systems, which is very difficult and costly to achieve. Recognizing when unexpected variation has changed the outcome of either training or assessment using a simulator is the responsibility of experienced human proctors and trainers. Simulated environments remain an approximation of the real world, but they also contain a level of complexity, which makes them as difficult to control as it is to control the real world.

Roger Smith, PhD
Florida Hospital Nicholson
Center, Celebration FL
roger.smith@flhosp.org

REFERENCES

1. Lampotang S. Unlike history, should a simulator not repeat itself? *Simul Healthc* 2015;10(6): 331–335.
2. Fujimoto R. *Parallel and Distributed Simulation Systems.* New York: John Wiley & Sons; 2000.
3. Kuhl F, Weatherly R, Dahmann J. *Creating Computer Simulation Systems: An Introduction to the High Level Architecture.* Upper Saddle River, NJ: Prentice Hall PTR; 1999.
4. Yilmaz L, Oren T. Toward replicability-aware modeling and simulation: changing the conduct of M&S in the information age. In: Tolk A, ed. *Ontology, Epistemology, and Teleology for Modeling and Simulation: Philosophical Foundations for Intelligent M&S Applications.* Heidelberg, Germany: Springer-Verlag; 2013.
5. Law A, Kelton W. *Simulation Modeling and Analysis.* 2nd ed. New York, NY: McGraw-Hill; 1991.
6. Steinman J. A unified approach to parallel simulation. *Proc of the 6th Workshop on Parallel and Distributed Simulation* 1992;6:75–84.

Surg Endosc (2015) 29:972–983
DOI 10.1007/s00464-014-3748-7

DYNAMIC MANUSCRIPTS

Comparative analysis of the functionality of simulators of the da Vinci surgical robot

Roger Smith · Mireille Truong · Manuela Perez

Received: 14 November 2013 / Accepted: 10 July 2014 / Published online: 15 August 2014

Abstract

Background The implementation of robotic technology in minimally invasive surgery has led to the need to develop more efficient and effective training methods, as well as assessment and skill maintenance tools for surgical education. Multiple simulators and procedures are available for educational and training purposes. A need for comparative evaluations of these simulators exists to aid users in selecting an appropriate device for their purposes.

Methods We conducted an objective review and comparison of the design and capabilities of all dedicated simulators of the da Vinci robot, the da Vinci Skill Simulator (DVSS) (Intuitive Surgical Inc., Sunnyvale, CA, USA), dV-Trainer (dVT) (Mimic Technologies Inc., Seattle, WA, USA), and Robotic Surgery Simulator (RoSS) (Simulated Surgical Skills, LLC, Williamsville, NY, USA). This provides base specifications of the hardware and software, with an emphasis on the training capabilities of each system.

Results Each simulator contains a large number of training exercises, DVSS = 40, dVT = 65, and RoSS = 52 for skills development. All three offer 3D visual images but use different display technologies. The DVSS leverages the real robotic surgeon's console to provide visualization, hand controls, and foot pedals. The dVT and RoSS created simulated versions of all of these control systems. They include systems management services which allow instructors to collect, export, and analyze the scores of students using the simulators.

Conclusions This study is the first to provide comparative information of the three simulators functional capabilities with an emphasis on their educational skills. They offer unique advantages and capabilities in training robotic surgeons. Each device has been the subject of multiple validation experiments which have been published in the literature. But those do not provide specific details on the capabilities of the simulators which are necessary for an understanding sufficient to select the one best suited for an organization's needs.

Keywords Robotic surgery · Robotic simulator · Training · Education · Comparative analysis

For every complex and expensive system, there emerges a need for training devices and scenarios which will assist new learners in mastering the use of the device and understanding how to apply it with value. In laparoscopic surgery, simulators have played an important role in improving the practice of surgery over the last 20 years [1, 2]. The same trends and values will likely apply to robotic surgery with the increased use of robotic technology for a growing variety of minimally invasive surgical procedures. The complexity, criticality, and cost associated with the effective application of the da Vinci surgical robot have stimulated the commercial creation of simulators, which replicate the operations of this robot. The objective of this paper was to provide

Electronic supplementary material The online version of this article (doi:10.1007/s00464-014-3748-7) contains supplementary material, which is available to authorized users.

R. Smith (✉)
Florida Hospital Nicholson Center, 404 Celebration Place, Celebration, FL 34747, USA
e-mail: roger.smith@flhosp.org

M. Truong
Florida Hospital Celebration Health, Celebration, FL 34747, USA

M. Perez
IADI-INSERM U947, Lorraine University, 54500 Nancy, France

Fig. 1 Simulators of the da Vinci surgical robot

comparative data on the functionality of the three commercially available robotic simulators as shown in Fig. 1:

- da Vinci Skill Simulator (Intuitive Surgical Inc., Sunnyvale, CA, USA);
- dV-Trainer (Mimic Technologies, Inc., Seattle, WA, USA); and
- RoSS (Simulated Surgical Skills LLC, Williamsville, NY, USA).

Each of these possesses unique traits which make them valuable solutions for different types of users and learning environments. This report is on the first of a three part comparative analysis of these devices. The first examines the functionality of each of the simulators and illustrates these capabilities side-by-side for ease of evaluation by potential users of each device. The second is a subjective usability evaluation of the simulators on similar exercises by novice (medical students), intermediate (residents and fellows), and expert (attending surgeons) subjects. The third is measure of the degree to which each simulator improves the actual robotic skills of a subject who is engaged in a two months training program with the device. This paper presents the results of the first study defining the functionality of the devices.

Materials and methods

Our department purchased each of the simulator devices which are being evaluated in these studies. This allowed us to objectively evaluate and comment on each device without undue influence from the manufacturers. Each simulator company was aware of the comparison project and provided information on their device in response to queries by our researchers, as noted below. We began by reviewing the users' manuals for the devices to collect details about each system [3–5]. We then interviewed representatives of each of the manufacturing companies for additional functional details. Finally, we performed our own experiments with each device to identify important comparative features across all devices.

We conducted a systematic literature review of all three simulators. The PubMed database of medical research was searched for all references to the devices through March 2013. References from retrieved articles were reviewed to broaden the search. The data extracted from these studies include training exercise modules, scoring systems, costs, educational impact, and validation methods. We identified 45 studies investigating simulation in robotic surgery.

Finally, we submitted our comparative data on the systems to the manufacturers of the devices to verify the accuracy of the information. Each company verified that the data presented in this analysis were accurate.

Results

Each of these devices is manufactured by a different company and provides a unique hardware and software solution for training and surgical rehearsal. The general features and capabilities of each are summarized in Table 1.

Table 1 Robotic simulator feature comparison

Features	DVSS	dV-Trainer	RoSS
System manufacturer	Intuitive Surgical Inc.	Mimic Technologies Inc.	Simulated Surgical Systems LLC
Specifications (simulator only)	Depth 7″ Height 25″ Width 23″ 120 or 240 V power	Depth 36″ Height 26″ Width 44″ 120 or 240 V power	Depth 44″ Height 77″ Width 45″ 120 or 240 V power
Specifications (complete system as shown in Fig. 1)	Depth 41″ Height 65″ Width 40″ 120 or 240 V power	Depth 36″ Height 59″ Width 54″ 120 or 240 V power	Depth 44″ Height 77″ Width 45″ 120 or 240 V power
Visual resolution	VGA 1,024 × 768	VGA 1,024 × 768	VGA 640 × 480
Components	Customized computer attached to da Vinci surgical console	Standard PC, visual system with hand controls, foot pedals	Single integrated custom simulation device
Support equipment	da Vinci Si surgical console, custom data cable	Adjustable table, touch screen monitor, keyboard, mouse, protective cover, custom shipping container	USB adapter, keyboard, mouse
Exercises	40 simulation exercises	65 simulation exercises	52 simulation exercises
Optional software	PC-based simulation management	Mshare curriculum sharing web site	Video and haptics-based procedure exercises (HoST)
Scoring method	Scaled 0–100 % with passing thresholds in multiple skill areas	Proficiency-based point system with passing thresholds in multiple skill areas	Point system with passing thresholds in multiple skill areas
Student data management	Custom control application for external PC. Export via USB memory stick	Export student data to delimited data file and graphical reports	Export student data to delimited data file
Curriculum customization	None	Select any combination of exercises. Set passing thresholds and conditions	Select specifically grouped exercises. Set passing thresholds
Administrator functions	Create student accounts on external PC. Import via USB memory stick	Create student accounts. Customize curriculum	Create student accounts. Customize curriculum
System setup	None	Calibrate controls	Calibrate controls
System security	Student account ID and password	PC password, Administrator password, Student account ID, and password	PC password, Administrator password, student account ID, and password
Simulator base price	$85,000	$99,200	$126,000
Support equipment price	$500,000	$9,800	$0
Total functional price	$585,000	$109,000	$126,000

Data are for simulator configurations available as of December 2013

Features and capabilities

Da Vinci Skill Simulator (DVSS) (Intuitive Surgical Inc.)

The DVSS consists of a customized computer package that attaches to the back of the surgeon's console of an actual da Vinci Si robot. This simulator connects to the surgeon's console via a single fiber optic networking cable identical to that used to connect the components of the actual robotic surgical system.

Advantages

Attached simulators of this type are usually referred to as "embedded trainers" because they take advantage of the equipment that has already been constructed, purchased, and installed for the use of the real system. These kinds of simulators are especially common in military facilities which face limited space and weight constraints. They can significantly reduce the hardware that must be purchased solely for simulation purposes. The U.S. Navy uses these kinds of

simulators aboard ships to reduce weight and space requirements, enabling them to train, while the ship is at sea.

Another significant advantage of an attached simulator is that it allows the trainee to use the actual controls from the real system to drive the simulator. This ensures that the training experience is almost identical in feel to the real system, which can contribute to higher transfer of skills from the training sessions to the real system. Additionally, this minimizes the amount of time spent for learning the unique functionalities of the simulator device and allows the trainee to focus the majority of his/her learning experience on skills acquisition and proficiency development. Finally, there is the cost advantage for the simulator device itself. Because much of the hardware and software expenses are already embedded in the real system, the simulator can be very economical to purchase.

Disadvantages

Attached simulators like the DVSS also come with inherent disadvantages to balance their positive traits.

The largest drawback is the availability and accessibility of a simulator which requires the real robotic system. An attached DVSS simulator cannot be used without access to an actual surgeon's console and therefore is only functional when the robotic system is not in surgery. This implies that the trainee would only be able to use the simulator outside of normal operating room working hours and would need logistical access to the robot and the simulator. da Vinci robots are expensive devices and hospitals typically attempt to maximize use of in order to recoup their investment. In a very active surgical hospital, it can be difficult to obtain access to a surgeon's console to support training with this simulator.

The DVSS is designed to connect to the surgeon's console using the same networking cable that connects the major robotic components. This makes the attachment and set-up process very easy for clinicians to master. However, it also means that the DVSS can only be used with the Si model surgeon's console. The previous S and Standard models use a different set of cables, which are not compatible with the simulator.

Similar to the military's experience with embedded and attached simulators, heavy usage of the DVSS comes with a corresponding heavy use of the surgeon's console. The Army and Navy have discovered that these types of simulators put more usage hours on real equipment controls which lead to more maintenance costs for those devices. Given the possibility of regular and continuous simulation training with such device, in addition to actual surgical usage, the real equipment may experience usage rates that are many times higher than normal for the equipment. Since the da Vinci systems operate under a maintenance contract that covers most service costs, the additional costs of maintenance are not born by the hospital owner but by the equipment vendor. The primary impact to the owner would only be in availability for both real surgeries and training events due to increased maintenance.

dV-Trainer (Mimic Technologies Inc.)

The dV-Trainer is a separate, stand-alone simulator of the da Vinci robot. The surgeon's console, controls, and vision cart are mimicked in hardware, while a 3D software model replicates the functions of the robotic arms and the surgical space.

Mimic Technologies also developed the core simulator software for the DVSS and used the same package in version 1.0 of their own dV-Trainer. As a result, the exercises in the DVSS and version 1.0 of the dV-Trainer are nearly identical. The current version 2.2 of the dV-Trainer has a number of new exercises which are not found in the DVSS, and the graphics have been upgraded so the visual presentation is no longer identical. The differences in visual presentation can be seen in Fig. 3 and 4.

The dV-Trainer consists of three major pieces of equipment and a number of smaller support pieces. The largest pieces are the "Phantom" hood which replicates the vision and hand controls of the da Vinci surgeon's console, the foot pedals of the surgeon's console, and a high-performance desktop computer which generates the 3D images and calculates the interactions with the surgeon's controls. Smaller support equipment includes a touch screen monitor, keyboard, and mouse to enable an instructor to guide the student through exercises and allow an administrator to manage the data that are collected.

Because the dV-Trainer replicates both the hardware and software of the da Vinci robot, it is a much larger system than the DVSS alone, though smaller than a real surgeon's console with the DVSS attached. It has the advantage of providing a training system that is completely independent of the need for any piece of the real surgical robot. The simulator can be configured to imitate either the S or the Si model of the da Vinci robot.

The disadvantage of this kind of system is that the simulated hardware is different than the real equipment and does not exactly replicate the feel of the real robotic equipment. The dV-Trainer uses its own unique hand controls which are connected to three cables for measuring movement, rather than the more precise arms that are used in the da Vinci robot. The dV-Trainer foot pedals look and function almost identically to the robotic foot pedals.

Robotic Surgery Simulator (Simulated Surgical Systems LLC)

The RoSS is also a complete, stand-alone simulator of the da Vinci robot. This device is designed as a single piece of

hardware that has a similar appearance to the surgeon's console of the robot. The hardware device includes a single 3D computer monitor, hand controls that are modified commercial force feedback devices, pedals that replicate either the S or the Si model of the da Vinci robot, and an external monitor for the instructor. The simulator can be configured to imitate either the S or the Si model of the da Vinci robot.

The hand controls are modified SensAble Omni Phantom™, force feedback, 3D space controllers (3D Systems Inc., Rock Hill, SC, USA). These devices have a much smaller range of motion than the controllers on the da Vinci robot, so require more frequent clutching than the actual robot. The 3D image is generated by a single computer monitor with polarized glasses, which generates a visual scene with less depth of field than the actual robot.

The company has developed a set of 3D virtual exercises that are unique from those found in both of the other simulators. They also provide optional video-based surgical exercises, called HoST modules, in which the user is guided through the movements necessary to complete an actual surgical procedure. At this writing, these modules are available for radical prostatectomy, hysterectomy, and cystectomy. These guided videos take advantage of the force feedback capabilities of the hand controllers to push and pull the student's hands to follow the simulated instruments on the screen. They require the student to perform specific movements accurately during the video before the operation will proceed.

Exercise modules

Each simulator allows an administrator or instructor to manage and organize student performance according to unique login credentials for the student. Alternatively, they all have a universal "guest" account to make the system accessible to anyone but without the ability to uniquely identify and track the performance of a specific student.

Once logged into each system, the instructor or the student navigates the instructional materials using the menu systems illustrated in Fig. 2. Since the intuitive skills simulator (DVSS) and the Mimic dV-Trainer provide very similar exercises and organizations, the navigation through the exercises is similar in form, though different in visual appearance. The RoSS simulator uses a very unique arced orbital menu for progressing through exercises.

Each simulator provides on-system instructions for every exercise in the form of textual documents and video demonstrations with spoken audible instructions.

Dvss

The DVSS contains 40 exercises organized into nine categories (Table 2). These begin with introductory video and audio instructions on how to use the robotic equipment and move through progressively more difficult skills (Table 3).

To prepare the student for success in each exercise, the simulator offers written instructions on the objective of each exercise prior to performance. There is also a narrated video of an instructor performing the exercise while explaining the necessary steps.

Upon completion of each exercise, the system automatically proceeds to a scoreboard showing the student's performance on the exercise. Details on the scoring systems of each simulator are discussed later in the article.

Figure 3 presents screenshots of some of the key exercises in the simulator. These include the Peg Board, Ring Walk, Energy Dissection, and Interrupted Suturing exercises. The suturing exercises on this simulator were developed by Simbionix USA Inc. (Cleveland, OH) for integration into the DVSS. This expansion of the system demonstrates the ability of the simulator platform to blend together exercises and scoring systems created by multiple independent vendors.

dV-Trainer

Most of the simulation software for Intuitive's DVSS was developed by Mimic Technologies. Therefore, version 1.0 of the DVSS and the dV-Trainer contained nearly identical exercises, closely matching menu systems, and identical scoring mechanisms. However, over time the two sets of software have diverged, and the current versions of the simulators differ in functionality and appearance. The current version of the dV-Trainer (v 2.2) contains 65 exercises organized into ten categories.

Though many of the exercises are identical between the DVSS and the dV-Trainer, the graphics resolution and details have been improved in version 2.2 of the dV-Trainer software. Since this system is driven by a commercial PC, which can easily be upgraded, it is possible for the hardware and software to evolve as newer computer technologies are available.

Just as with the DVSS, the dV-Trainer simulator offers written instructions on the objective of each exercise prior to performance. There is also a narrated video of an instructor performing the exercise while explaining the necessary steps. Upon completion of each exercise, the system automatically proceeds to a scoreboard showing the student's performance on the exercise.

Figure 4 presents screenshots of some of the key exercises in the dV-Trainer simulator. These include the Peg Board, Match Board, Tubal Anastomosis, and Energy Switching exercises.

RoSS

The RoSS simulator contains 52 unique exercises, organized into five categories, and arranged from introductory

Fig. 2 Comparative simulator exercise menus

Table 2 DVSS exercise categories

Surgeon console overview	An introduction to the controls of the da Vinci robot
Endowrist manipulation 1	Basic hand movements and usage of the wristed instruments
Camera and clutching	Basic foot clutching for both the camera and the third arm
Endowrist manipulation 2	Intermediate use of the hands and wristed instruments
Energy and dissection	Use of the energy pedals and associated instruments
Needle control	Focused exercises for dexterous manipulation of a curved surgical needle
Needle driving	Repetitive exercises for needle driving
Games	Challenging and entertaining game environments to apply the skills learned
Suturing skills	Suturing exercises with needle, following suture, knot-tying, and tissue closure

to more advanced (Table 4), just as in the other two simulators. The RoSS system of exercises is unique in that they list fewer exercises but provide three different difficulty levels for most of them where each level is actually a unique exercise.

Similar to the other simulators, the RoSS includes a narrated video showing an instructor performing the exercise. Upon completion of an exercise, the simulator automatically proceeds to the scoreboard for the exercise.

The RoSS contains a unique capability that is not found in either of the other simulators called “Hands-on Surgical Training” or “HoST.” This is an integration of surgical skills exercises with a video of an actual surgery. Videos of actual surgical procedures play in the surgeon’s visual space, overlaid with animated icons, which instruct the student to perform specific actions during the progression of the surgery video.

Table 3 dV-Trainer exercise categories

Surgeon console overview	An introduction to the controls of the da Vinci robot
Endowrist manipulation	Basic and intermediate use of the hand controllers and wristed instruments
Camera and clutching	Basic foot clutching for both the camera and the third arm
Energy and dissection	Use of the energy pedals and associated instruments
Needle control	Focused exercises for dexterous manipulation of a curved surgical needle
Needle driving	Repetitive exercises for needle driving
Troubleshooting	Introduction to error recovery on the da Vinci robot
Games	Challenging and entertaining game environments to apply the skills learned
Suturing skills	Suturing exercises with needle, following suture, knot-tying, and tissue closure
RTN	VR exercises specifically build to match physical devices in use by the research training network of sites led by Lehigh Valley Hospital

The necessary actions are prompted with audio instructions. For the HoST exercise to progress, the student must perform the specific actions at specific times. The simulator will pause the video and allow the student to repeat the action until it is performed as required by the instructions.

The hand controllers of the RoSS simulator are modified versions of a commercially available 3D haptic input device called the Omni Phantom™. This product uses internal motors and gears to apply haptic feedback to the hand movements of the user. For the HoST exercises, the simulator uses this capability to move the student’s hands in sync with the movements of the surgeon’s instruments in the master video.

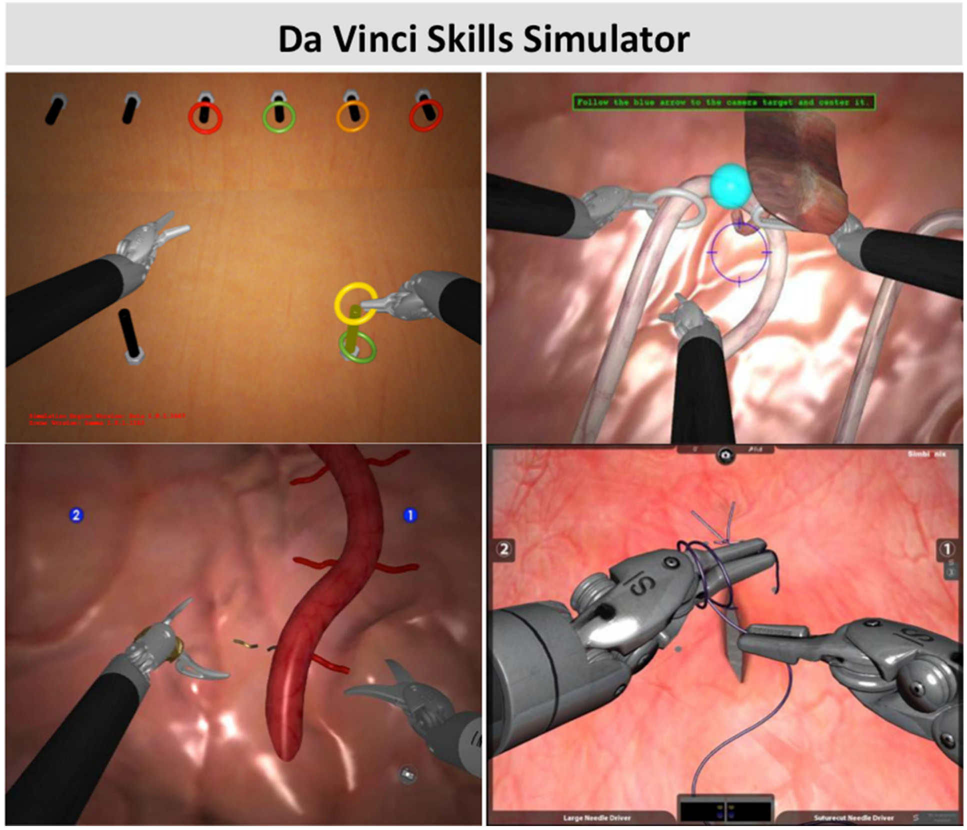


Fig. 3 Selected DVSS exercise images

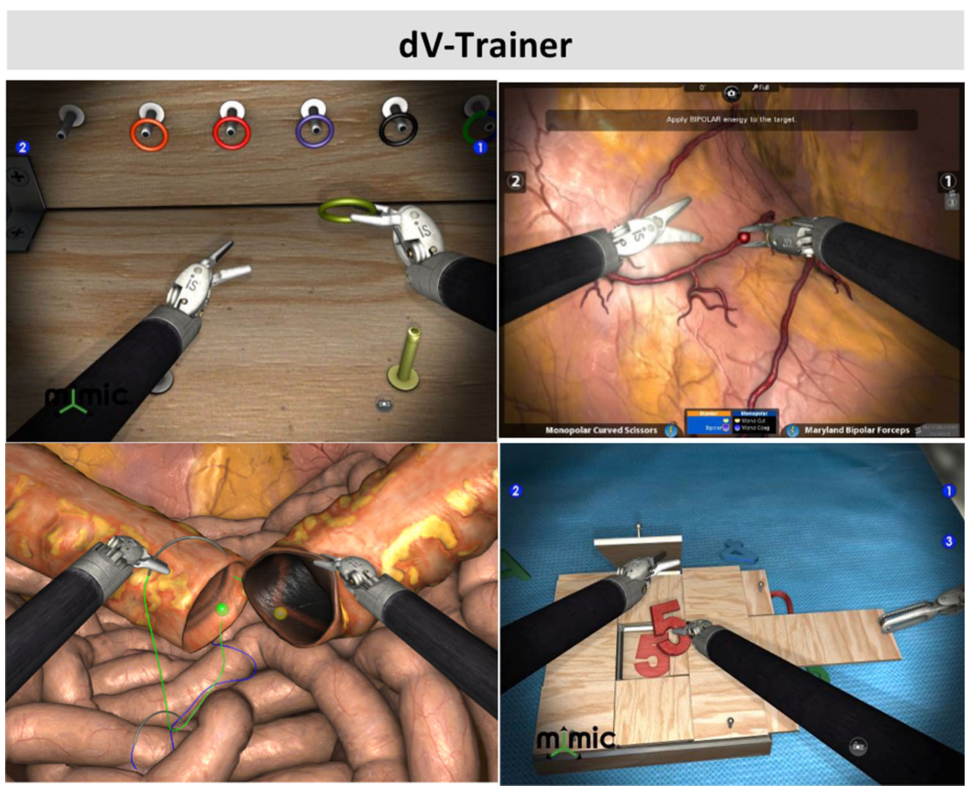


Fig. 4 Selected dV-Trainer exercise images

Table 4 RoSS exercise categories

Orientation module	Introduction to the surgeon controls of the da Vinci robot
Motor skills	Development of precise controls of the instruments, including spatial awareness
Basic surgical skills	Instruction on handling a needle, using electrocautery pedals and instruments, and the use of scissors on the robot
Intermediate surgical skills	Control of the fourth arm, blunt tissue dissection, and vessel dissection
Hands-on surgical training	Video and haptic-guided instruction through specific surgical procedures

Figure 5 provides screenshots of the motor skills ball placement, intermediate vessel dissection, 4th arm tissue retraction, and HoST radical prostatectomy.

Proficiency scoring system

Each of the three simulators provides a different scoring method. All three use the host computer to collect data on the performance of the student at the controls in multiple performance areas. With this data, they provide a score for specific performance traits, as well as combining all of these into a single composite score of performance for the entire exercise. The algorithm used to create this composite score is described in the user's manuals of each of the simulators. Examples of each of these scoreboards are shown in Fig. 6.

In addition to the objective metrics that can be collected by the computer, the developers of each simulator have been challenged to provide thresholds, which indicate whether the student's score is considered a "passing" or "failing" performance. All three have identified threshold scores, which would indicate acceptable and warning scoring levels. These are commonly interpreted as "passing" (above acceptable threshold) and "failing" (below warning threshold), with a "warning" area between the two thresholds. These thresholds create green, yellow, and red performance areas, which can be used to visually communicate the quality of the student's performance in each area of measurement. Each simulator also provides a single composite score for the entire exercise.

Each of the simulators gives the student a single overall score for performance on an exercise. To achieve this, an algorithm was needed to combine very different types of metrics. For example, the number of seconds to complete an exercise needs to be combined with milliliters of blood loss, centimeters of instrument movement, number of instrument collisions, and other similarly varied metrics. As in most educational environments, this is achieved by converting each metric into a score, which falls between some defined minimum and maximum value. Most people understand this concept from their academic experience in which all assignments were graded in the range from 0 to 100 % or between 0 points and the maximum total points for all assignments. These normalizations make it possible to create a single composite score of the student's

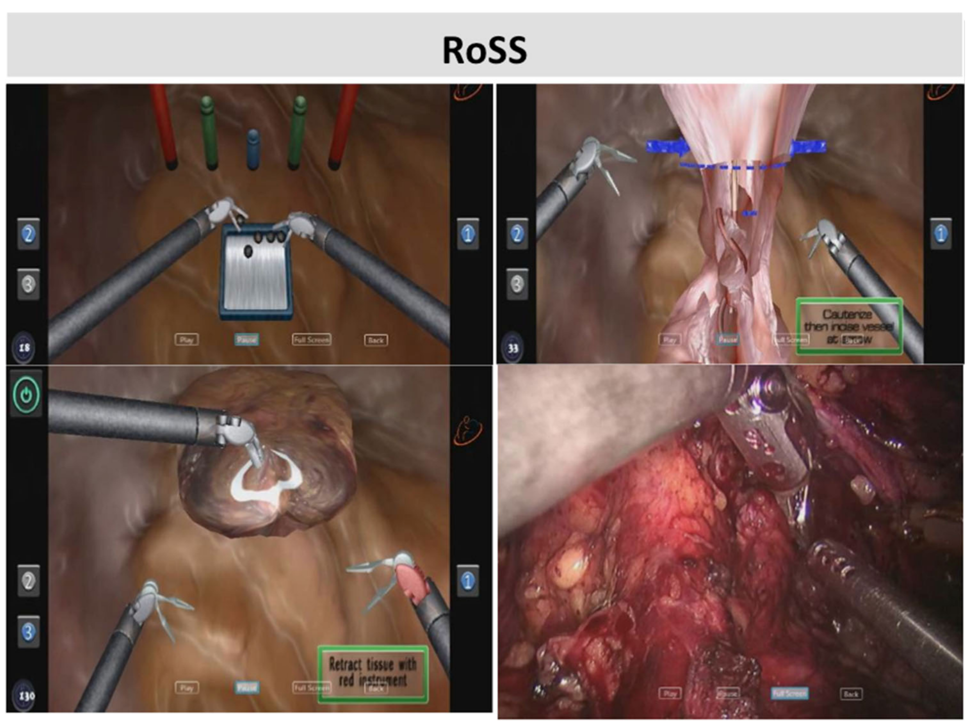


Fig. 5 Selected RoSS exercise images

Fig. 6 Example scoreboards from each simulator

performance across multiple assignments. This same approach has been used in the simulators, where the resulting composite metric may be a total point score or a percentage.

The simulator manufacturers work with robotic surgeons to establish the relative values of each measure used in the composite score, just as they did for the threshold levels described earlier. Because these evaluations are the opinions of the individuals who collaborated with the company on the development of the system. The dV-Trainer and the RoSS both provide the ability for a system administrator to adjust these levels to meet the needs of unique curriculum, courses, and students being evaluated.

Table 5 DVSS and dV-Trainer scoring method

Overall score	Composite evaluation of the exercise performance
Time to complete	Number of seconds to complete the exercise
Economy of motion	Number of centimeters of instrument tip movement
Instrument collisions	Number of times that the instruments touched each other
Excessive instrument force	Number of seconds that excessive robotic force was applied against objects in the environment
Instrument out of view	Number of centimeters that an instrument tip moved outside of the viewing area
Master workspace range	Radius in centimeters than contains the movement of the instrument tips
Drops	Number of objects dropped from the grasp of the instruments

DVSS

The DVSS performance scoring method has a number of metrics which are applied to every exercise, and others which are only used for exercises in which they are relevant. Table 5 presents the metrics which are applicable to all exercises. For details on the more specialized metrics, the reader may consult the user's manual for the simulator.

Because the DVSS is a closed, turnkey system with an ease of use similar to the actual surgical robot, most of the data displays, and threshold adjustments found in the other simulators are not available in this device. Simulator settings are determined by the manufacturer and cannot be changed by the user.

dV-Trainer

Originally, the DVSS and the dV-Trainer shared the same scoring method, but more recent versions of the dV-Trainer offer both this original "version 1.0" scoring method, as well as a new "version 2.0" method based on the proficiency measured from experienced surgeons. The skills measured are the same (Table 3), but the interpretation of those into a score is different. The instructor can select the preferred scoring method for each curriculum that is constructed in the dV-Trainer. The newer scoring method uses total points earned rather than percentages. The passing and warning thresholds can be adjusted by the administrator.

RoSS

The principles behind the scoring system on the RoSS are the same as those for the DVSS and the dV-Trainer. However, most of the metrics collected are different. The standard measurements are shown in Table 6.

Like each of the other simulators, there are multiple displays of the performance data for a student. The initial display presented at the completion of an exercise shows a horizontal bar, which is colored green, yellow, or red to indicate passing or failing. The magnitude of the bar is a rough measure of the quality of performance (Fig. 6). Additional displays show the numeric score and its relative position to a passing threshold.

Table 6 RoSS scoring method

Overall score	Composite evaluation of the exercise performance
Camera usage	Optimal movement of camera
Left tool grasp	Optimal number of tool grasps with left hand tool
Left tool out of view	Distance left hand tool is out of view
Number of errors	Number of collision or drop errors in an exercise
Right tool grasp	Optimal number of tool grasps with right hand tool
Right tool out of view	Distance right hand tool is out of view
Time	Time to complete the exercise
Tissue damage	Number of times that instruments damaged tissue with excessive force or unnecessary touches
Tool–Tool collision	Number of times tools touched each other

System administration

All of the simulators contain system configuration and student management functions, which require a special administrator account to access and modify. These allow instructors to create curriculum and scoring methods, which are unique to the lessons they are offering. They also allow an instructor or administrator to create new student accounts and export student scores for evaluation and analysis outside of the simulator device. Some course instructors use this capability to create custom performance reports for students who attend the courses.

DVSS

For the DVSS, most of the administrator functionality is fixed within the delivered system. The administrator can create specific user profiles for the simulator using a dedicated program on a separate external PC. This program, the "DVSS Manger", allows the administrator to create a profile for the user. The profile can then be loaded onto a USB memory stick and inserted into the USB port on the DVSS. The simulator will automatically read this data in and display the user names at the login screen.

Similarly, the USB memory stick can be inserted into the DVSS, and the performance data collected from exercises performed by each user will be automatically loaded onto the USB stick. This stick can then be inserted in the PC, and the data will be loaded into the management software on the external PC and exported to a delimited file for formatting and analysis in a spreadsheet program.

The entire transfer process is automated and the contents of the USB stick are completely erased and reloaded each time. The stick cannot safely be used for any purpose other than as the transfer mechanism between the two devices. This method is meant to create an ease of use similar to the real robot.

dV-Trainer

The administrator on a dV-Trainer has the ability to create new user accounts, specify S or Si representation, create new curriculum, set passing thresholds, and export user data for analysis.

The simulator contains 65 exercises, any combination of which can be organized into a curriculum for a specific course. The administrator creates the new curriculum name and then adds each exercise that should be part of the curriculum. This set of exercises can be organized into phases or folders to match the course that is being taught. For example, an instructor may have a curriculum that consists of a warm-up with easy exercises, pre-course evaluations, and post-course evaluations. These would appear as three separate sections within the curriculum.

The administrator can export data from the simulator according to multiple criteria. The export may include all of the data on the machine, or subsets defined by the unique user ID, date range, completion status, or a specific exercise.

The capabilities provided for an administrator of the dV-Trainer are significantly more robust than those available on the other two simulators.

RoSS

The RoSS administrator account is used to create student accounts. Each user can then be assigned a specific subset of the entire simulator curriculum.

For the RoSS system, the administrator can assign portions of the curriculum hierarchy, which are applicable to a specific user. The curriculum is organized such that customization consists of selective subsets of the hierarchy of exercises, rather than the ability to select specific exercises in unique combinations.

The administrator can also edit the passing thresholds for each exercise. This allows a site to create curriculum, which is considered passing for practitioners at different levels, such as medical students, residents, attending, and specialists.

The scores can be exported as individual delimited data files for each student account. These can then be removed from the system for analysis and recording.

Validation of devices

Validation studies serve to determine whether a simulator can actually teach or assess what it is intended to teach or assess. In medical simulation, there are generally accepted

Table 7 Validation of robotic surgical simulators

Validation	DVSS	dV-Trainer	RoSS
Face: subjective realism of the simulator	Hung [7] Kelly [8] Liss [9]	Lendvay [10] Kenney [11] Sethi [12] Perrenot [13] Korets [14] Lee [15] Schreuder [16]	Seixas-Mikelus [17] Stegemann, [18]
Content: judgment of appropriateness as a teaching modality	Hung [7] Hung [19] Kelly [8] Liss [9]	Kenney [11] Sethi [12] Perrenot [13] Lee [15]	Seixas-Mikelus [17] Colaco, [20]
Construct: able to distinguish experienced from inexperienced surgeon	Hung [7] Kelly [8] Liss [9] Finnegan [21]	Kenney [11] Perrenot [13] Korets [14] Lee [15] Schreuder [16] Connolly [22] Lendvay [23]	Raza [24]
Concurrent: extent to which simulator correlates with "gold standard"	Hung [19] Tergas [25]	Perrenot [13] Korets [14] Lee [15] Lerner [26]	Chowriappa, [27]
Predictive: extent to which simulator predicts future performance	Hung [19] Tergas [25] Culligan [28]		

validity classifications, which include face, content, construct, concurrent, and predictive validity [6]. Face and content validity are considered subjective approaches, while the other three are objective approaches to validation.

Table 7 provides a summary of the published validation studies for these simulators. All three have publications establishing face, content, construct, and concurrent validation. Only published studies investigate the predictive validity of the DVSS [19, 25, 28]. Recent presentations also explore the validity of the RoSS curriculum [29] and the RoSS' HoST procedural modules [30].

Conclusions

Simulators play an important role in providing a training experience and a platform for evaluation of novices who are trying to master complex skills in many fields. When a task is simple, consequences for failure are minimal, and equipment is inexpensive, there is little motivation for creating a dedicated simulation device. However, when the task to be mastered is complex, there is a need for a device that can objectively measure the performance of the trainee and provide feedback that leads to improved performance. When the consequences of a mistake can be lethal, there is a need for a safe environment in which to develop expertise without threatening the wellbeing of others. When equipment or disposables are expensive to use, there is a need for a tool that can provide at least entry-level familiarization and skill development without undue financial demands. All three of these conditions are characteristic of the process for learning robotic surgery. So it is not surprising that market forces have led to the creation of multiple simulators of the robotic system and the skills to use it.

This article represents the first part of a comprehensive analysis of robotic surgical simulators. The second part is a subjective opinion survey on the usability of the simulators. Subjects for this survey will include attending surgeons, fellows, residents, and medical students without prior experience using the simulation devices. The third part will include a select group of surgical fellows who will participate in a two-month experiment in which each practices on one of the simulators, while their performance is measured every 2 weeks to assess for changes and maintenance of skill levels. The experiment is designed to determine which simulator has the greatest positive impact on robotic surgical performance, and the degree to which those improvements are retained across a period of inactivity.

The three simulators described in this article are complex systems, which are significantly less costly than the actual da Vinci robotic surgical system and can be operated at a fraction of the cost of the instruments required by this robot. Furthermore, da Vinci robots are predominantly used for daily surgery, decreasing their availability for training. There are currently no available studies directly comparing the three simulators, and therefore until those studies are performed, no universal recommendation can be made for one device over the other, and a decision to use one simulator over the other should be based on unique and individual needs.

Acknowledgments The authors wish to thank the following representatives of each simulator company for their assistance in collecting data and images, as well as verifying the accuracy of the contents of the paper. Peter Dominick, Intuitive Surgical Inc.; Jeff Berkley, Mimic Technologies, Inc.; and Kesh Kesavadas, Simulated Surgical Systems, LLC.

Funding Source U.S. Army Telemedicine and Advanced Technology Research Center. Grant #: W81XWH-11-2-0158.

Disclosures Florida Hospital has received a research grant from Mimic Technologies Inc. for a project unrelated to this study. Roger Smith, Mireille Truong and Manuela Perez have no conflicts of interest or financial ties to disclose.

References

1. Schout BM, Hendriks AJ, Scheele F et al (2010) Validation and implementation of surgical simulators: a critical review of present, past, and future. J Surg Endosc 24(3):536–546
2. Wohaibi EM, Bush RW, Earle DB et al (2010) Surgical resident performance on a virtual reality simulator correlates with operating room performance. J Surg Res 160(1):67–72
3. Skills simulator for the da Vinci Si surgical system (users manual) (2012) Intuitive Surgical Inc., Sunnyvale
4. dV-Trainer robotic simulator users manual (2012) Mimic Technologies Inc., Seattle
5. Robotic surgery simulator users manual: for models S and Si (2012) Simulated Surgical Systems LLC, Williamsville
6. McDougall EM (2007) Validation of surgical simulators. J Endourol 21(3):244–247
7. Hung AJ, Zehnder P, Patil MB (2011) Face, content and construct validity of a novel robotic surgery simulator. J Urol 186(3):1019–1024
8. Kelly DC, Margules AC, Kundavaram CR (2012) Face, content, and construct validation of the da Vinci Skills Simulator. J Urol 79(5):1068–1072
9. Liss MA, Abdelshehid C, Quach S (2012) Validation, correlation, and comparison of the da Vinci trainer(™) and the da Vinci surgical skills simulator(™) using the mimic(™) software for urologic robotic surgical education. J Endourol 26(12):1629–1634
10. Lendvay, Casale P, Sweet R, Peters C (2008) Initial validation of a virtual-reality robotic simulator. J Robot Surg 2(3):145–149
11. Kenney P, Wszolek M, Gould J, Libertino J, Moinzadeh A (2009) Face, content, and construct validity of dV-Trainer, a novel virtual reality simulator for robotic surgery. J Urol 73(6):1288–1292
12. Sethi AS, Peine WJ, Mohammadi Y (2009) Validation of a novel virtual reality robotic simulator. J Endourol 23(3):503–508
13. Perrenot C, Perez M, Tran N (2012) The virtual reality simulator dV-Trainer(®) is a valid assessment tool for robotic surgical skills. J Surg Endosc 26(9):2587–2593
14. Korets R, Mues AC, Graversen JA, Gupta M, Benson MC, Cooper KL, Landman J, Badani KK (2011) Validating the use of the mimic dV-Trainer for robotic surgery skill acquisition among urology residents. J Urol 78(6):1326–1330
15. Lee JY, Mucksavage P, Kerbl DC (2012) Validation study of a virtual reality robotic simulator–role as an assessment tool? J Urol 187(3):998–1002
16. Schreuder HWR, Persson JEU, Wolswijk RGH, Ihse I, Schijven MP, Verheijen RHM (2014) Validation of a novel virtual reality simulator for robotic surgery. Sci World J 2014:507076. doi:10.1155/2014/507076
17. Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G (2010) Face validation of a novel robotic surgical simulator. J Urol 76(2):357–360
18. Stegemann AP, Kesavadas T, Rehman S, Sharif M, Rao A, DuPont N, Shi Y, Wilding,G, Hassett J, Guru K (2012) Development, implementation and validation of a simulation-based curriculum for robot-assisted surgery. Presented at the AUA poster session, Atlanta, 19–23 May 2012
19. Hung AJ, Patil MB, Zehnder P (2012) Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. J Urol 187(2):630–637
20. Colaco M, Balica A, Su D (2012) Initial experiences with RoSS surgical simulator in residency training: a validity and model analysis. J Robot Surg 7(1):71–75
21. Finnegan KT, Meraney AM, Staff I (2012) da Vinci skills simulator construct validation study: correlation of prior robotic experience with overall score and time score simulator performance. J Urol 80(2):330–335
22. Connolly M, Seligman J, Kastenmeier A, Goldblatt M, Gould JC (2014) Validation of a virtual reality-based robotic surgical skills curriculum. Surg Endosc. doi:10.1007/s00464-013-3373-x
23. Lendvay TS, Brand TC, White L, Kowalewski T, Jonnadula S, Mercer LD, Khorsand D, Andros J, Hannaford B, Satava RM (2013) Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study. J Am Coll Surg 216:1181–1192
24. Raza SJ, Chowriappa A, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody J, Menon M, Kesavadas T, Guru KA. (2013) Construct validation of the fundamental skills of robotic surgery (FSRS) curriculum. Podium presentation, AUA annual congress, San Diego
25. Tergas AI, Sheth SB, Green IC, Giuntoli RL 2nd, Winder AD, Fader AN (2013) A pilot study of surgical training using a virtual robotic surgery simulator. JSLS 17:219–226
26. Lerner M, Ayalew M, Peine W, Sundaram C (2010) Does training on a virtual reality robotic simulator improve performance on the da Vinci® surgical system? J Endourol 24(3):467–472
27. Chowriappa A, Raza SJ, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody JO, Menon M, Kesavadas T, Guru KA (2013) Development & validation of a composite grading & scoring system for robot-assisted surgical training—the robotic skills assessment scale. Podium presentation, AUA annual congress, San Diego
28. Culligan P, Gurshumov E, Lewis C, Priestley J, Komar J, Salamon C (2014) Predictive validity of a training protocol using a robotic surgery simulator. Female Pelvic Med Reconstr Surg 20:48–51
29. Stegemann A, Kesavadas T, Rehman S, Sharif M, Rao M, duPont N, Shi Y, Wilding G, Hassett J, Guru K (2013) Development, implementation, and validation of a simulation-based curriculum for robot-assisted surgery. J Urol 81(4):767–774
30. Ahmed K, Chowriappa A, Din R, Field E, Raza SJ, Fazili A, Samarasekera D, Kaouk J, Eichel L, Joseph J, Ghazi A, Mohler J, Peabody J, Eun D, Shi Y, Wilding G, Kesavadas T, Guru KA (2013) A multi-institutional randomized controlled trial of an augmented-reality based technical skill acquisition module for robot-assisted urethro-vesical anastomosis. Podium presentation, AUA annual congress, San Diego

Surg Endosc (2016) 30:3720–3729
DOI 10.1007/s00464-015-4667-y

Robotic surgery simulation validity and usability comparative analysis

Alyssa Tanaka[1] · Courtney Graddy[2] · Khara Simpson[3] · Manuela Perez[4] · Mireille Truong[3] · Roger Smith[1]

Received: 4 March 2015 / Accepted: 6 November 2015 / Published online: 18 November 2015

Abstract

Background The introduction of simulation into minimally invasive robotic surgery is relatively recent and has seen rapid advancement; therefore, a need exists to develop training curriculums and identify systems that will be most effective at training surgical skills. Several simulators have been introduced to support these aims—the daVinci skills simulator, Mimic dV-Trainer, Surgical Simulated Systems' RoSS, and Simbionix Robotix Mentor. While multiple studies have been conducted to demonstrate the validity of these systems, studies comparing the perceived value of these devices as tools for education and skills are lacking.

Methods Subjects who qualified as medical students or physicians ($n = 105$) were assigned a specific order to use each of the three simulators. After completing a demographic questionnaire, participants performed one exercise on the three simulators and completed a second questionnaire regarding their experience with the device. After using all systems, they completed a final questionnaire, which detailed their comparative preferences. The subject's performance metrics were also collected from each simulator.

Results The data confirmed the face, content, and construct validity for the dV-trainer and skills simulator. Similar validities could not be confirmed for the RoSS. >80 % of the time, participants chose the skills simulator in terms of physical comfort, ergonomics, and overall choice. However, only 55 % thought the skills simulator was worth the cost of the equipment. The dV-Trainer had the highest cost preference scores with 71 % of respondents feeling it was worth the investment.

Conclusions Usability can affect the consistency and commitment of users of robotic surgical simulators. In a previous study, these simulators were objectively reviewed and compared in terms of their system capabilities. Collectively, this work will offer end-users and potential buyers a comparison of the perceived value and preferences of robotic simulators.

Keywords Simulation · Validation · Robotic surgery · Training · Usability

Medicine has come to the conclusion that the Halstedian training model (i.e., See one, do one, teach one) is no longer sufficient for teaching complex skills, particularly robotic surgical skills [1]. With the introduction of robotic technology between patient and surgeon, a need to master new skills has emerged. A number of virtual reality simulators have been developed to support the training and acquisition of such skills. Currently, the commercially available robotic simulators include: the da Vinci skills simulator (dVSS) by Intuitive Surgical Inc., also known as the "Backpack Simulator"; the dV-Trainer from Mimic Technologies Inc.; the RoSS by Simulated Surgical Sciences LLC; and the Robotix Mentor from Simbionix (Fig. 1). All of these da Vinci simulators utilize a visual scene that is presented in a computer-generated 3D environment providing challenging tests for practicing dexterity and machine operations. Originally, the simulated

✉ Alyssa Tanaka
Alyssa.tanaka@flhosp.org

1 Florida Hospital Nicholson Center, 404 Celebration Place, Celebration, FL 34747, USA

2 Florida Hospital Celebration Health, Celebration, FL, USA

3 Columbia University Medical Center, New York, NY, USA

4 University of Lorraine-Nancy, Nancy, FR, USA

Fig. 1 Simulators of the da Vinci robotic surgical system

exercises trained basic robotic skills; however, with advances in technology, surgeons can now train for specific procedures (e.g., partial nephrectomy and hysterectomy).

The work described in this paper is the second part of a three-phase analysis to study the effectiveness of these simulators and applications to the education of robotic surgeons. In the first phase, the authors evaluated and compared the objective characteristics of three simulators (dVSS, dV-Trainer, and RoSS). The Simbionix Robotix Mentor was not included because it was under development at the time of this research. This analysis provided a head-to-head comparison of the systems and found that they varied greatly in their hardware and software.

In the dVSS, the trainee operates the simulated environment using the actual da Vinci surgical console. The simulator is a custom computer, appended to the surgical console through the surgical data port. While the simulator costs approximately $85,000, the surgical console costs $500,000 incurring an investment of $585,000. Using this simulator, users can train with the actual hardware they would use during surgery; however, this requires availability of the surgical console, which may be fully scheduled in the operating room. Few hospitals have a dedicated training console, meaning that users do not have ready access to the simulator. The second system is a standalone system that utilizes a high-performance graphic/gaming computer, connected to a custom desktop viewing and control device that replicates the hardware of the da Vinci surgeon's console. This system shares similar software with the dVSS, but does not require the use of actual da Vinci hardware. The cost of this simulator is approximately $96,000. The third system is composed of a completely customized replica of the da Vinci surgeon's console. Internally the simulator contains a graphic computer, a 3D viewing system, and commercial Omni Phantom haptic controllers. This simulator uses unique software and costs approximately $126,000 [2].

The validity of medical and surgical simulators is typically evaluated using the categories defined by McDougal [3]. This paper defines the most commonly recognized forms of validation as: *face, content, construct, concurrent, and predictive validity. Face validity* is typically assessed informally by users and indicates whether the simulator is an accurate representation of the actual system (i.e., the realism of the simulator). *Content validity* is the measure of the appropriateness of the system as a teaching modality. Experts who are knowledgeable about the device typically assess this via a formal evaluation. *Construct validity* is the ability of a simulator to measure what it is intended to measure. Often this is characterized by the simulator's ability to differentiate between users' experience level. *Concurrent validity* is the extent to which the simulator correlates with the "gold standard" for training, and *predictive validity* is the extent to which the simulator can predict a user's future surgical performance. Collectively, concurrent and predictive validity are known as criterion validity and are used as measures of the simulator's ability to correlate trainee performance with their real-life performance. Face and content validity are most effective in evaluating the ability of a simulator to train a surgeon; however construct, concurrent, and predictive validity are most useful for evaluating the effectiveness of a simulator to assess a trainee.

The validity of all three simulators has been examined separately (Table 1), and to our knowledge, there is no comparative research of all three systems. The current study therefore compares the three commercially available da Vinci simulators and details the findings for face, content, and construct validity of these systems. The purpose of this is to provide end-users and potential buyers with a head-to-head evaluation of the value and usability of the systems.

Materials and methods

Participants in this study included medical students, residents, fellows, and attending physicians. Participants were recruited from the University of Central Florida College of Medicine, courses held at the Florida Hospital Nicholson Center, and two surgical robotics conferences (World Robotics Gynecology Congress and Society of Robotic

Table 1 da Vinci simulator validation studies from Smith et al. [2]

Validation	DVSS	dV-Trainer	RoSS
Face: subjective realism of the simulator	Hung [4] Kelly [5] Liss [6]	Lendvay [7] Kenney [8] Sethi [9] Perrenot [10] Korets [11] Lee [12] Schreuder [13]	Seixas-Mikelus [14] Stegemann, [15]
Content: judgment of appropriateness as a teaching modality	Hung [4] Hung [20] Kelly [5] Liss [6]	Kenney [8] Sethi [9] Perrenot [10] Lee [12]	Seixas-Mikelus [14] Colaco [17]
Construct: Able to distinguish experienced from inexperienced surgeon	Hung [4] Kelly [5] Liss [6] Finnegan [18]	Kenney [8] Perrenot [10] Korets [11] Lee [12] Schreuder [13] Connolly [19] Lendvay [20]	Raza [21]
Concurrent: Extent to which simulator correlates with "gold standard"	Hung [16] Tergas [22]	Perrenot [10] Korets [11] Lee [12] Lerner [23]	Chowriappa, [24]
Predictive: Extent to which simulator predicts future performance	Hung [16] Tergas [22] Culligan [25]		

Surgeons Scientific Meeting). Subjects were excluded from participating if they had participated in a formal robotic simulation-training course to eliminate preference bias. Each participant was categorized into one of three groups (i.e., expert, intermediate, or novice) according to the self-reported number of robotic cases performed. Individuals who had performed 0–19 robotic cases were categorized as novices, individuals with 20–99 robotic cases were considered to be intermediates, and individuals with 100 or more cases were considered to be experts.

After being categorized into an experience level, each participant was assigned a specific order in which they used each of the simulators (Fig. 2). This alternating order was implemented to identify and eliminate any potential bias that may exist by using a specific system first. All participants completed one exercise on each of the simulators. The tasks chosen were Peg Board 1 in both the dV-Trainer and the dVSS and Ball Placement 1 in the RoSS. The same task was used for both the dV-Trainer and the dVSS because these systems share similar software and exercises. The RoSS software contains unique exercises, and Ball Placement 1 was chosen because it trains the same basic skills as Peg Board 1.

After completing the exercise on a simulator, participants completed a post-questionnaire (Survey 1), which asked for feedback regarding their experience on that specific simulator. After using all three systems, subjects completed a second post-questionnaire (Survey 2), which asked them to compare all three systems to each other. The participant's performance metrics were also collected from each of the simulators.

Results

The novice group ($n = 37$) had performed an average of 2 robotic cases, the intermediate group ($n = 31$) on average performed 54 cases, and the expert group ($n = 37$) performed 336 cases. Sixty-two percent of subjects were men, and 38 % were women with an average age of 43. On average, participants had 15 years in practice and 3 years of robotic experience. Seventy-six percent were attending physicians, and 73 % of participants were currently or had received robotic surgery training, while 41 % provided that they train residents and fellows. A one-way ANOVA verified a difference in the average age and number of years in

Participant	Simulator Order
Novice 01	CBA
Novice 02	BAC
Novice 03	ACB
Novice 04	CBA
Intermediate 01	CBA
Intermediate 02	BAC
Intermediate 03	ACB
Intermediate 04	CBA
Expert 01	CBA
Expert 02	BAC
Expert 03	ACB
Expert 04	CBA

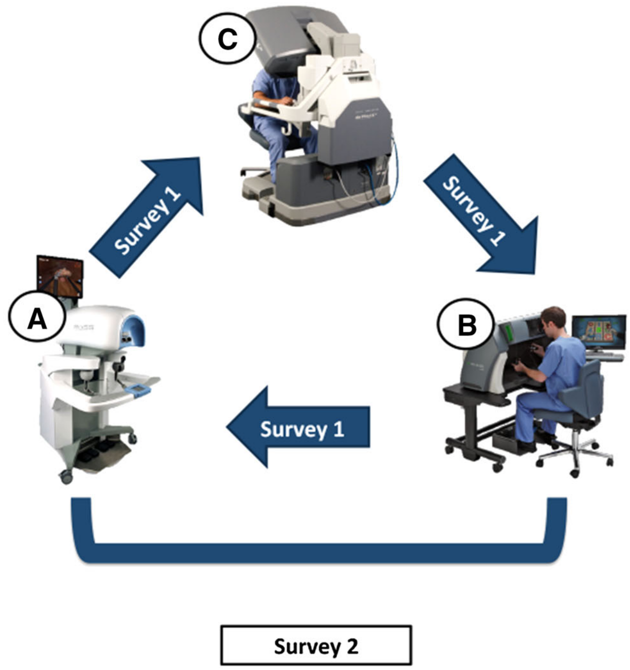


Fig. 2 Example of rotating order and research process

practice of participants based on the classification of expert, intermediate or novice (number of robotic procedures). This is to be expected since higher ages typically imply a higher number of years of practice and resultant larger numbers of robotic procedures.

The types of validity evaluated in this experiment were face, content, and construct. To analyze the systems for face validity and content validity, questions from Survey 1 were used. The questions were evaluated on a five-point Likert scale (i.e., Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, and Strongly Agree). As recommended by Van Nortwick et al. [26], face validity was analyzed by expert and intermediate feedback only as these are the users most familiar with the robotic system; however, only expert feedback was used for content validity because they have the best ability to judge the appropriateness of the system as a training tool. For construct validity, performance metrics such as overall score, time to complete, number of errors, and economy of motion were analyzed (Table 2). Specifically, time and economy of motion were chosen due to a previous study by Perrenot et al. [10] indicating that these are highly relevant indicators of expertise in robotic surgery.

Face validity

A Chi-square test of independence was used to evaluate the distribution of scores for a specific simulator in relation to the order of the system's presentation to the subject. This analysis indicated that there was no difference in participants' responses according to the order in which the systems were presented; and established that no bias was present due to the presentation order ($p > 0.05$). These questions asked participants to evaluate whether the hand controllers on the simulator were effective for working in the simulated environment (Question 1) and if the device is a sufficiently accurate representation of the real robotic system (Question 4). For both questions, the RoSS had the lowest average score, dV-Trainer had the second highest score, and the dVSS had the highest score of the three (Table 3). A repeated measures ANOVA verified that the answers were statistically different for both questions ($p < 0.001$).

Content validity

As seen in Table 4, 100 % of participants either agreed or strongly agreed that the 3D graphical exercises in the dVSS were effective for teaching robotic skills, while 59 % disagreed or strongly disagreed that the RoSS' capabilities were effective. When asked if the scoring system effectively communicated their performance, 88 % of dVSS users agreed or strongly agreed, while 79 % of dV-Trainer users agreed or strongly agreed. Similarly, 91 and 82 % of participants agreed or strongly agreed that the dVSS and dV-Trainer, respectively, effectively guided them to improve their performance, while only 36 % felt the RoSS provided the same guidance.

Table 2 Description of data used for types of validity

Type of validity	Evaluation	Type of participant	Question/metric
Face validity	Survey 1	Expert and intermediate	Q1: The hand controllers on this simulator are effective for working in the simulated environment (Likert)
			Q4: The device is a sufficiently accurate representation of the real robotic system (Likert)
Content validity	Survey 1	Expert	Q2: The 3D graphical exercises in the simulator are effective for teaching robotic skills (Likert)
			Q5: The scoring system effectively communicates my performance on the exercise (Likert)
			Q6: The scoring system effectively guides me to improve performance on the simulator (Likert)
Construct validity	Simulator	Experts and novices	Overall score (points)
			Number of errors (count)
			Time to complete (seconds)
			Economy of motion (centimeters)

Table 3 Mean scores from a 5-point Likert scale on face validity

Face validity ($n = 68$)	DVSS	dV-Trainer	RoSS
Q1: The hand controllers on this simulator are effective for working in the simulated environment.	4.80	3.62	2.17
Q4: The device is a sufficiently accurate representation of the real robotic system.	4.65	3.45	1.82

Table 4 Percentages of Likert responses for content validity questions

Content validity ($n = 34$)					
Likert Score	Strong dis (%)	Disagree (%)	Neither (%)	Agree (%)	Strong agree (%)
Q2: The 3D graphical exercises in the simulator are effective for teaching robotic skills.					
DVSS	0	0	0	35.3	64.7
dV-Trainer	2.9	5.9	11.8	50.0	29.4
RoSS	20.6	38.2	17.6	17.6	5.9
Q5: The scoring system effectively communicates my performance on the exercise					
DVSS	2.9	5.9	2.9	38.2	50.0
dV-Trainer	2.9	2.9	14.7	55.9	23.5
RoSS	17.6	20.6	26.5	29.4	5.9
Q6: The scoring system effectively guides me to improve performance on the simulator.					
DVSS	0	0	8.8	61.8	29.4
dV-Trainer	2.9	2.9	11.8	61.8	20.6
RoSS	18.2	18.2	27.3	33.3	3.0

Construct validity

The overall score, number of errors, time to complete, and economy of motion scores collected by the simulators for experts ($n = 37$) and novices ($n = 37$) were used to compare construct validity (Table 5). Intermediate subjects were not included in the construct validity analysis because it was only necessary to determine whether the simulator could distinguish specifically between novice and expert users. Overall score is synthesized from multiple metrics and is specific to the individual simulator. This metric was available in the dVSS and the dV-Trainer; however, the overall score metric is not automatically exported by the RoSS and therefore was not analyzed for this system. Instead, the number of errors was used for the RoSS. For all of the simulators, higher overall score values are better, while lower economy of motion, time, and number of error values are better preferred.

Table 5 Mann–Whitney *U* test level of significance on construct validity measures

	DVSS	dV-Trainer	RoSS
Time to complete	*$p < 0.001$*	*$p < 0.001$*	$p = 0.221$
Overall score	*$p < 0.01$*	$p = 0.061$	n/a
Economy of motion	$p = 0.216$	*$p < 0.001$*	$p = 0.566$
Number of errors	n/a	n/a	$p = 0.644$

Values italicized are statistical significance ($p < 0.05$)

For the RoSS, the analysis has 23 missing data points because the system does not report scores when a user exceeds a maximum exercise time or chooses to terminate the exercise before completion. This resulted in a sample of 30 experts and 21 novices on this system. A Mann–Whitney *U* test showed that the distributions of time ($p = 0.221$), number of errors ($p = 0.644$), and economy of motion ($p = 0.566$) were not statistically different for the experts compared to the novice group on this simulator.

The dV-Trainer analysis of experts ($n = 37$) and novices ($n = 37$) had three missing values for economy of motion and completion time and five for the overall score metric, thus, the analysis contained varying number of subjects. The distribution of the overall scores was not significantly different for the expert compared to the novice group ($p = 0.061$). These tests did confirm statistical differences for economy of motion ($p < 0.001$) and time to complete ($p < 0.001$), with a lower economy of motion value and shorter completion time for experts compared to novices.

The dVSS analysis included all novice ($n = 37$) and expert ($n = 37$) participants. Time to complete ($p < 0.001$) and overall score ($p = 0.006$) were significantly different for the expert compared to the novice group. The expert group had a higher overall score and a shorter completion time compared to the novice group. However, economy of motion did not show a statistical difference with this analysis ($p = 0.216$).

The relationship between experience and performance metrics was more specifically analyzed in terms of the self-reported number of cases of all participants ($n = 105$) using a nonparametric correlation coefficient (Spearman's). For the RoSS, 30 participants were excluded from the analysis. For the participants that were included in the analysis ($n = 75$), there was not a significant correlation between time to complete ($p = 0.181$), number of errors ($p = 0.563$), or economy of motion ($p = 0.390$) with the total number of robotic cases performed (Fig. 3).

For the dV-Trainer, four participants were excluded from the entire analysis and two participants were excluded from the overall score analysis (overall score $n = 99$; economy of motion and time to complete $n = 101$). The analysis verified a statistically significant correlation between overall score ($p = 0.03$), economy of motion ($p < 0.01$), and time to complete ($p < 0.01$). The correlation value was negative for economy of motion and time to complete, showing that with a greater number of robotic cases, the time taken and distance moved decreased. The correlation was positive for overall score indicating that the participants' score increased with the number of robotic cases performed (Fig. 4).

For the dVSS, two participants were excluded from the analysis ($n = 103$). A statistically significant difference was found between overall score ($p = 0.01$) and time to complete ($p < 0.01$). The correlation value was negative for time and positive for overall score, signifying that with more robotic cases the time taken decreased and the score increased. There was not a statistically significant correlation between economy of motion and the total number of robotic cases performed ($p = 0.105$) (Fig. 5).

Usability (preference)

The questions from Survey 2 were used to understand the preference of the subjects when using the simulators. All subjects were included in this analysis except for two participants who were dropped from the analysis because they did not complete the questionnaire. The participants' responses to the following usability questions can be seen in Fig. 6:

- If you are (were) a program director, which simulator would you choose for your trainees;
- In which simulator were you physically more comfortable;
- Which simulator had the best hand controls;
- Which simulator had the best foot controls;
- Which simulator had the best 3D vision;
- Were you feeling stressed or annoyed by any of the simulators?

Overall, most participants preferred the dVSS and indicated that they would choose this device as a training system if they were a program director. Participants not only felt most comfortable in the dVSS, but also felt that the system had the best control and vision equipment. The least preferred system was the RoSS, which most participants also agreed made them feel stressed or annoyed. Ten percent of participants also responded that they felt stressed or annoyed by both the dV-Trainer (dVT) and the RoSS.

Cost

All participants were also asked to provide feedback on their simulator preference in terms of the cost of the system. The responses were analyzed in terms of the frequency of the responses given. Most participants felt that the dV-Trainer was worth the investment, while most felt that the RoSS was not. When asked about the dVSS, only

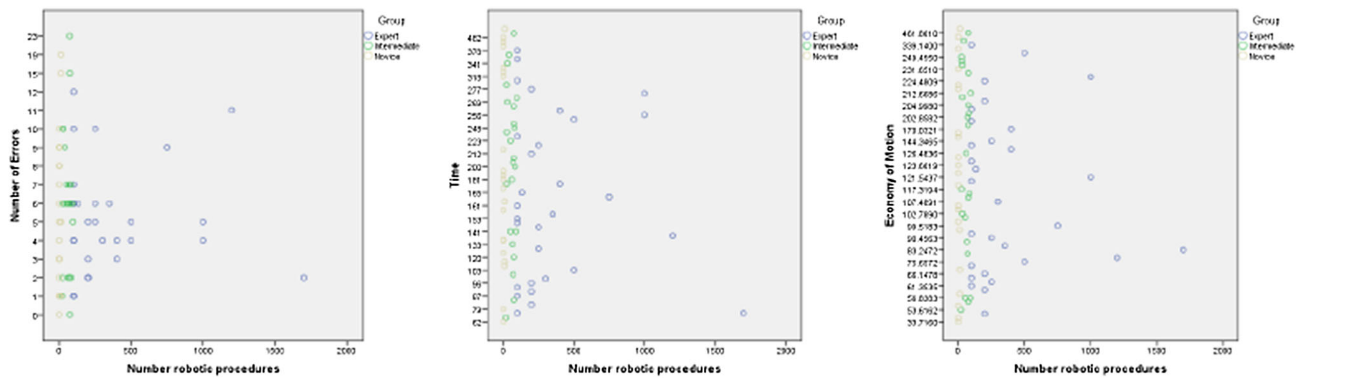


Fig. 3 Graphs of correlation between experience and metrics on the RoSS

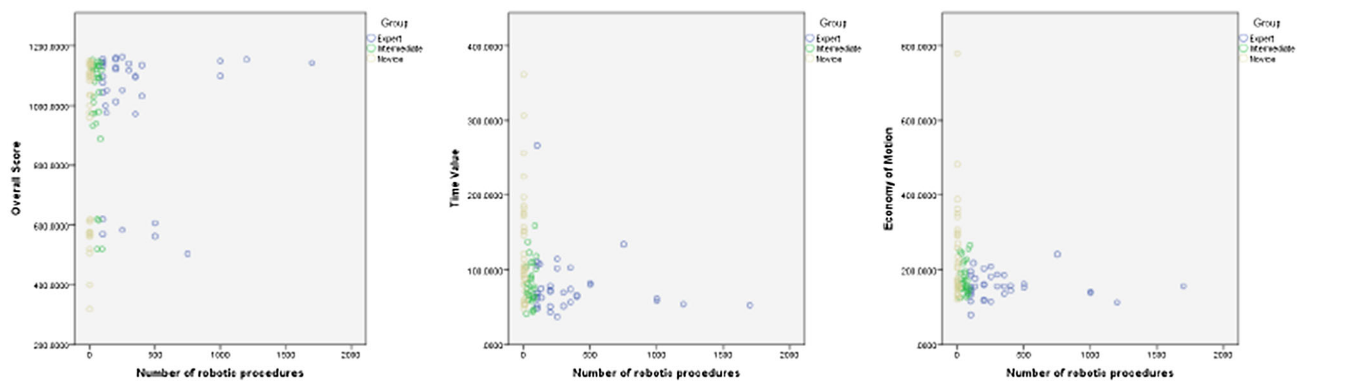


Fig. 4 Graphs of correlation between experience and metrics on the dV-trainer

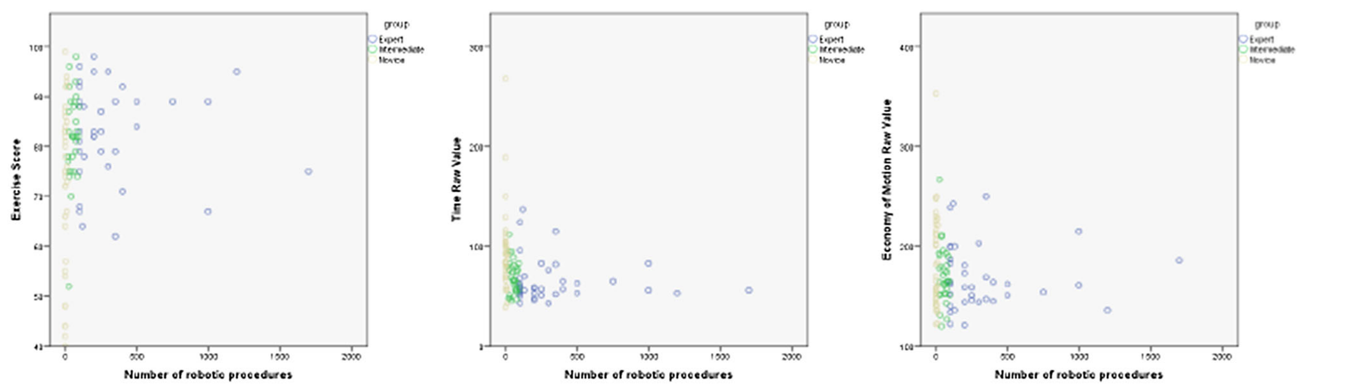


Fig. 5 Graphs of correlation between experience and metrics on the dVSS

56 % of participants agreed that it was worth the investment (Fig. 7).

Discussion

The aim of this study was to conduct a comparison of the three commercially available simulators used to train surgeons on the daVinci robotic surgical system. The study was performed to assist potential buyers in making a purchasing and deployment decision regarding robotic simulators. This study provides information about the face, content, and construct validity, as well as usability of the systems.

The simulators were perceived to be different in their representation of the real robotic system. The dVSS was most preferred in terms of ergonomics and usability; however, most participants did not feel that this system was

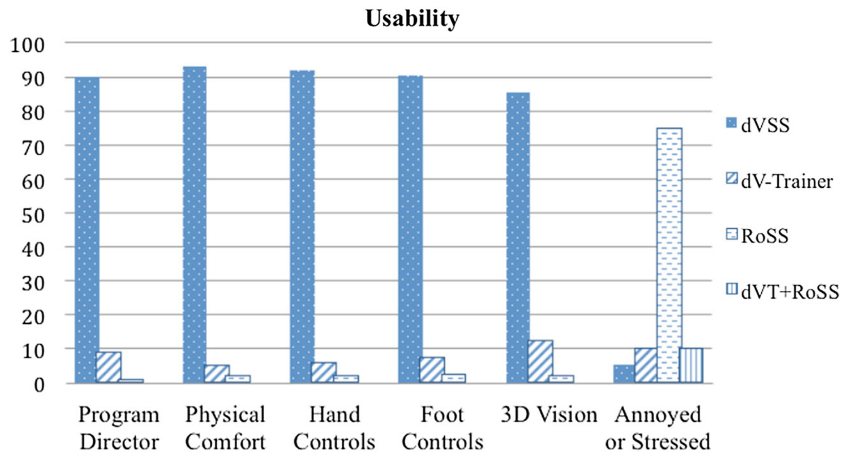


Fig. 6 Description of usability responses

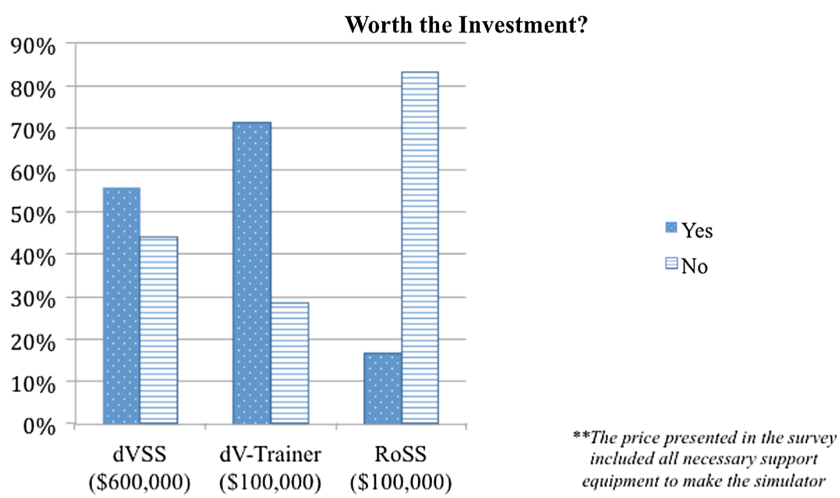


Fig. 7 Description of cost preferences

worth the investment. The costs provided in the questionnaire included all equipment needed to make the simulator functional. While the simulator itself only costs $85,000, it is impossible to use without the $500,000 da Vinci surgeon console. By leveraging the actual da Vinci hardware, this simulator allows for a more realistic experience, but limits the availability and creates a higher cost for training than other robotic simulators. Economy of motion was not able to differentiate novices from experts in the dVSS, which could be attributed to the ease of use of the controllers allowing novices to move the controls as efficiently as experts. The generous workspace of the dVSS could also have an impact on the lack of difference.

In terms of cost, most participants agreed that the dV-Trainer had the best cost-effectiveness. In contrast to the dVSS, the dV-Trainer is a standalone simulator and does not require the support of the daVinci hardware to operate. This allows for better accessibility and requires less of an investment for training. The overall score aspect of construct validity in the dV-Trainer may not have shown a difference between novices and experts due to the way that the scoring is developed. The scoring system is constructed with a "ceiling" that prevents users from achieving a high overall score without attaining high scores across multiple metrics.

The RoSS was the least preferred system for comfort and other usability aspects (i.e., hand controls, foot controls, and 3D interface), with most participants feeling stressed or annoyed when using the system. This study was unable to validate the face, content, or construct validity for this system. Currently, there is limited data available that confirms construct validity of the RoSS. Contrary to Raza [21], this study was unable to confirm a difference between experts and novices in terms of time taken to complete the exercise. As stated previously, time and economy of motion are considered highly relevant

measures of expertise levels [10] and should distinguish between these groups in the simulators.

To our knowledge this three-part study is the first to compare three of the available simulators. This study involved the largest sample size and diversity of participants (i.e., experience levels, number of robotic cases, and subspecialty type) thus far in relevant publications. The results from this research will help guide the choice of simulators used for future studies at Florida Hospital and may also influence decisions at other laboratories. However, a limitation to the study was the lack of consistency in the available exercises and scoring systems across the three systems. A consideration for future studies will be to use more complex exercises and increase the depth of the face and content validity evaluation. Future research should continue to critically evaluate surgical simulators, including new iterations of da Vinci simulators (e.g., the Simbionix Robotix Mentor). There is limited research on the transfer of skills from simulators to the actual da Vinci system. Future studies could investigate the transfer of training from a simulator to the surgical system via a dry laboratory assessment or actual procedure.

Acknowledgments The authors wish to thank the representatives of each simulator company for their assistance in collecting images: da Vinci skills simulator photos ©2013 Intuitive Surgical, Inc., used with permission; dV-Trainer Simulator photos ©2013 Mimic Technologies, Inc., used with permission; RoSS Simulator photos ©2013 Simulated Surgical Systems, LLC., used with permission; and Robotix Mentor photos ©2014 Simbionix USA, Corp., used with permission. This work is funded by U.S. Army Telemedicine and Advanced Technology Research Center. Grant #: W81XWH-11-2-0158.

Compliance with ethical standards

Disclosures Alyssa Tanaka, Courtney Graddy, Khara Simpson, Manuela Perez, Mireille Truong and Roger Smith have no conflicts of interest or financial ties to disclose.

References

1. Cameron JL (1997) William Stewart Halsted. Our surgical heritage. Ann Surg 225(5):445
2. Smith R, Truong M, Perez M (2014) Comparative analysis of the functionality of simulators of the da Vinci surgical robot. J Surg Endosc 29(4):972–983
3. McDougall EM (2007) Validation of surgical simulators. J Endourol 21(3):244–247
4. Hung AJ, Zehnder P, Patil MB (2011) Face, content and construct validity of a novel robotic surgery simulator. J Urol 186(3):1019–1024
5. Kelly DC, Margules AC, Kundavaram CR (2012) Face, content, and construct validation of the da Vinci Skills simulator. J Urol 79(5):1068–1072
6. Liss MA, Abdelshehid C, Quach S (2012) Validation, correlation, and comparison of the da Vinci Trainer(TM) and the da Vinci surgical skills simulator(TM) using the Mimic(TM) software for urologic robotic surgical education. J Endourol 26(12):1629–1634
7. Lendvay Casale P, Sweet R, Peters C (2008) Initial validation of a virtual-reality robotic simulator. J Robot Surg 2(3):145–149
8. Kenney P, Wszolek M, Gould J, Libertino J, Moinzadeh A (2009) Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. J Urol 73(6):1288–1292
9. Sethi AS, Peine WJ, Mohammadi Y (2009) Validation of a novel virtual reality robotic simulator. J Endourol 23(3):503–508
10. Perrenot C, Perez M, Tran N, Jehl JP, Felblinger J, Bresler L, Hubert J (2012) The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. J Surg Endosc 26(9):2587–2593
11. Korets R, Mues AC, Graversen JA, Gupta M, Benson MC, Cooper KL, Landman J, Badani KK (2011) Validating the use of the mimic dV-trainer for robotic surgery skill acquisition among urology residents. J Urol 78(6):1326–1330
12. Lee JY, Mucksavage P, Kerbl DC (2012) Validation study of a virtual reality robotic simulator—role as an assessment tool? J Urol 187(3):998–1002
13. Schreuder HWR, Persson JEU, Wolswijk RGH, Ihse I, Schijven MP, Verheijen RHM (2014) Validation of a novel virtual reality simulator for robotic surgery. Sci World J 2014:507076. doi:10.1155/2014/507076
14. Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G (2010) Face validation of a novel robotic surgical simulator. J Urol 76(2):357–360
15. Stegemann AP, Kesavadas T, Rehman S, Sharif M, Rao A, DuPont N, Shi Y, Wilding G, Hassett J, Guru K (2012) Development, implementation and validation of a simulation-based curriculum for robot-assisted surgery. Presented at the AUA poster session, May 19–23, 2012 Atlanta, GA
16. Hung AJ, Patil MB, Zehnder P (2012) Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. J Urol 187(2):630–637
17. Colaco M, Balica A, Su D (2012) Initial experiences with RoSS surgical simulator in residency training: a validity and model analysis. J Robot Surg 7(1):71–75
18. Finnegan KT, Meraney AM, Staff I (2012) da Vinci skills simulator construct validation study: correlation of prior robotic experience with overall score and time score simulator performance. J Urol 80(2):330–335
19. Connolly M, Seligman J, Kastenmeier A, Goldblatt M, Gould JC (2014) Validation of a virtual reality-based robotic surgical skills curriculum. Surg Endosc. doi:10.1007/s00464-013-3373-x
20. Lendvay TS, Brand TC, White L, Kowalewski T, Jonnadula S, Mercer LD, Khorsand D, Andros J, Hannaford B, Satava RM (2013) Virtual reality robotic surgery warm-up improves task performance in a dry laboratory environment: a prospective randomized controlled study. J Am Coll Surg 216:1181–1192
21. Raza SJ, Foghi S, Chowriappa A et al (2014) Construct validity of the key components of the fundamental skills of robotic surgery (FSRS) curriculum: a multi-institution prospective study. J Surg Educ 71(3):3126–3324
22. Tergas AI, Sheth SB, Green IC, Giuntoli RL 2nd, Winder AD, Fader AN (2013) A pilot study of surgical training using a virtual robotic surgery simulator. J Soc Laparoendosc Surg Soc Laparoendosc Surg 17:219–226
23. Lerner M, Ayalew M, Peine W, Sundaram C (2010) Does training on a virtual reality robotic simulator improve performance on the da Vinci® surgical system? J Endourol 24(3):467–472
24. Chowriappa A, Raza SJ, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody JO, Menon M, Kesavadas T, Guru KA (2013) Development and validation of a composite grading and scoring system for robot-assisted surgical training: the robotic skills assessment scale. Podium presentation, AUA annual congress, San Diego

25. Culligan P, Gurshumov E, Lewis C, Priestley J, Komar J, Salamon C (2014) Predictive validity of a training protocol using a robotic surgery simulator. Female Pelvic Med Reconstr Surg 20:48–51

26. Van Nortwick S, Lendvay T, Jensen A, Wright A, Horvath K, Kim S (2010) Methodologies for establishing validity in surgical simulation studies. Surgery 147(5):622–630

Using Simulators to Measure Communication Latency Effects in Robotic Telesurgery

Roger Smith, PhD
Florida Hospital Nicholson Center
Celebration, FL
roger.smith@flhosp.org

Sanket Chauhan, MD
University of Minnesota Medical School
Minneapolis, MN
sanket@umn.edu

ABSTRACT

Robotic surgical technology was originally developed by the US Army and DARPA as a tool to enable telesurgery at a distance. The Intuitive da Vinci system now provides a robotic surgical tool in a traditional operating room. But research continues into the extension of this capability to patients that are remote from the surgeon's location. In this paper we describe the interim results of experiments into the effects of communication latency in the safe execution of robotic telesurgeries. These experiments were carried out with the Mimic dV-Trainer, a simulator of the da Vinci robot, which was configured to insert defined levels of latency into the visual and command data streams between a surgeon and the operating field. Subjects were asked to perform four basic robotic surgical exercises. They were allowed to rehearse these in a zero latency environment and with a randomly assigned latency between 100ms and 1,000ms. Then each subject performed each exercise for measurement and analysis in our research.

This experiment measured the degradation of human surgical performance across a range of latency conditions. This paper reports on the comparison of the level of experience of the surgeons with their performance in a latency-effected environment. The data collected thus far refutes our hypothesis that more experienced surgeons would be more successful at managing the effects of latency and would perform better than those with less experience. Subjects in our experiment show no correlation between experience and successful performance under latency. The ability to manage latency in tele-operations may be shared between remote surgery and the control of a remotely piloted UAV's and UGV's. The results of our experiments may suggest that experience as a traditional pilot does not necessarily contribute to useful skills in flying UAV's or driving UGV's when latency is present.

ABOUT THE AUTHORS

Roger Smith, PhD, is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading technology implementation through the development of alliances with industry, the military, academic institutions, physician networks and governing medical associations. This includes identifying, executing and managing industry, military and federally funded simulation, modeling and training projects. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 11 book chapters, and over 100 journal and conference papers. His most recent book is *Innovation for Innovators: Leadership in a Changing World.* He has served on the editorial boards of *Transactions on Modeling and Computer Simulation* and *Research Technology Management*.

Sanket Chauhan, MD, is a Robotic Urology Fellow at the University of Minnesota Medical School. Prior to this he was with the Florida Hospital, Global Robotics Institute and an instructor of Urology at the University of Central Florida's College of Medicine. Dr. Chauhan's research interests include developing new technologies for the future of surgery, telesurgery, surgical education, advanced surgical technologies, surgical simulation and the use of virtual reality and augmented reality in surgery. He has published more than 25 papers in peer reviewed journals and has authored 3 book chapters. Dr Chauhan is committed to surgical education using next generation VR based simulators. He is a member of the program committee for International Association for Science and Technology for Development (IASTED) Robotics and Control conference in 2010, and the World Robotic Surgery Symposium.

Using Simulators to Measure Communication Latency Effects in Robotic Telesurgery

Roger Smith, PhD
Florida Hospital Nicholson Center
Celebration, FL
roger.smith@flhosp.org

Sanket Chauhan, MD
University of Minnesota Medical School
Minneapolis, MN
sanket@umn.edu

BACKGROUND

Robotic surgery has been the topic of science fiction and scientific research for decades. As early as 1942, Robert A. Heinlein published the story "Waldo" in *Astounding Science Fiction*. He described the use of gloves and a harness to allow Waldo Jones to control mechanical arms of any size from large industrial and construction equipment to miniature tools for electronic and surgical work. The Industrial Revolution gave us many of the tools needed to extend the capabilities of the human body, but the Information Age gave us the computerized control systems necessary to effectively manipulate these devices. Surgical robots are a marriage of mechanical, electrical, optical, and software systems that can empower a human surgeon to peer into a patient's body with magnified stereo vision, probe the internal organs, and perform effective surgery without fully opening the patient's body.

In 1985, the PUMA 560 was used to accurately place a needle for a brain biopsy using CT guidance (Kwoh et al, 1988). In 1988, the PROBOT at Imperial College London, was used to perform prostate surgery. In 1992, Integrated Surgical Systems introduced ROBODOC to mill precise fittings in the femur for hip replacement. Intuitive Surgical leveraged the research work of the Defense Advanced Research Projects Agency (DARPA) and used those technologies to create the da Vinci Surgical System which they introduced in 1997. Computer Motion followed a similar path and fielded the AESOP and ZEUS robotic systems (Figure 1), which were later acquired by Intuitive Surgical (Satava, 1998; FDA, 2005).

Figure 1. ZEUS Surgical Research Robot

Intuitive Surgical's da Vinci robot is currently the only FDA approved device for robotic surgery on human patients. This system senses the surgeon's hand movements and translates them into scaled-down micro-movements to manipulate tiny instruments inside the body. It also detects and filters out any tremors in the hand movements, so that they are not expressed robotically. The camera used in the system provides a true stereoscopic picture transmitted to and viewed through a surgeon's console (Figure 2).

These devices opened the door for the realization of surgery-at-a-distance, a.k.a. telesurgery, in which a surgeon is able to extend his reach and perform surgical procedures at a significant distance from the patient. This capability has been demonstrated under unique conditions by multiple experiments (Himpens, 1998; Janetschek, 1998; Fabrizio, 2000; Sterbis, 2007). Our research project at the Florida Hospital Nicholson Center is demonstrating the maturity of the existing telecommunication infrastructure within a hospital system to support daily, on-demand telesurgery right now. Our experiments are based on the da Vinci surgical robot (Intuitive Surgical, Inc.) and the dV-Trainer simulator (Mimic Technologies, Inc.).

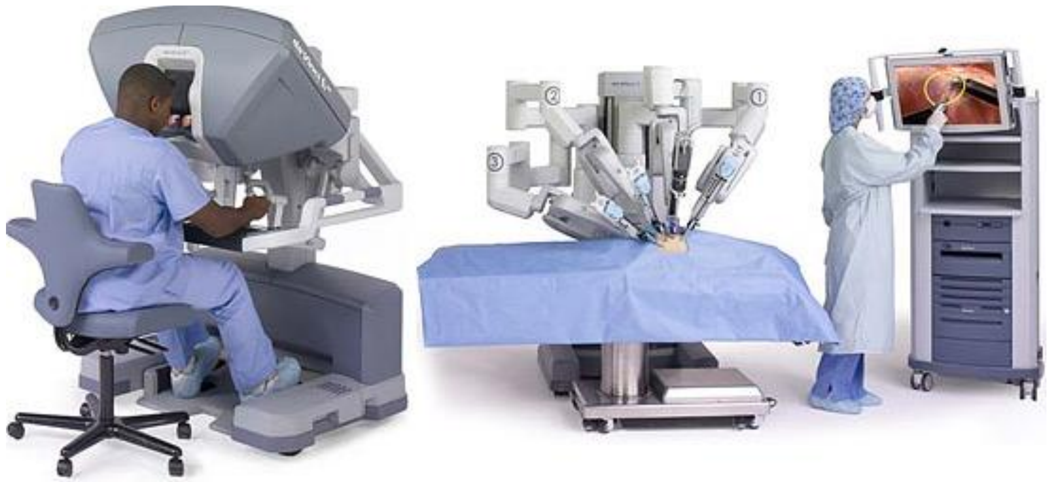


Figure 2. da Vinci Surgical Robot (Intuitive Surgical, Inc.)

METHODS

We explore the effects of communication latency on surgeon performance. This latency effect is created using the dV-Trainer simulator (Figure 3) of the da Vinci surgical robot (Hung, 2011; Kennedy 2009). The simulator allows the insertion of specific levels of controlled latency so that the user's physical movements are not manifest by the simulated instruments until after the defined latency period has elapsed.

Figure 3. dV-Trainer Simulator (Mimic Technologies, Inc.)

During actual telesurgery, the messages sent between the surgeon's machine and the remote patient station will be delayed due to the speed of light and the message routing that occurs on the internet. Determining how much latency can be safely tolerated in surgery is an important question (Anvari, 2005 and 2007). This experiment hypothesizes that there are two distinct thresholds of performance under increasing latency. The first is the level of latency at which a surgeon can first detect that his or her movements are being affected by the communication link. Any communication latency lower than this level is imperceptible and potentially non-invasive to the surgical procedure. Hence, if such levels can be achieved in the real world, then telesurgery may be safe for human surgery right now. The second level is the point at which the surgeon's performance is degraded to the point that the surgery cannot be performed safely (Marescaux, 2002; Lum, 2009). This level is identified through both simulator measured performance and the expert opinion of the surgeon. Between the first and second thresholds, a surgeon may be able to successfully control the effects of latency and perform a safe and successful procedure. Beyond the second threshold, telesurgery would be considered unsafe with the available equipment (Figure 4).

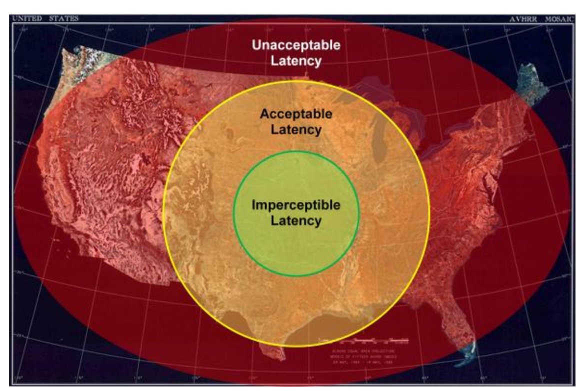
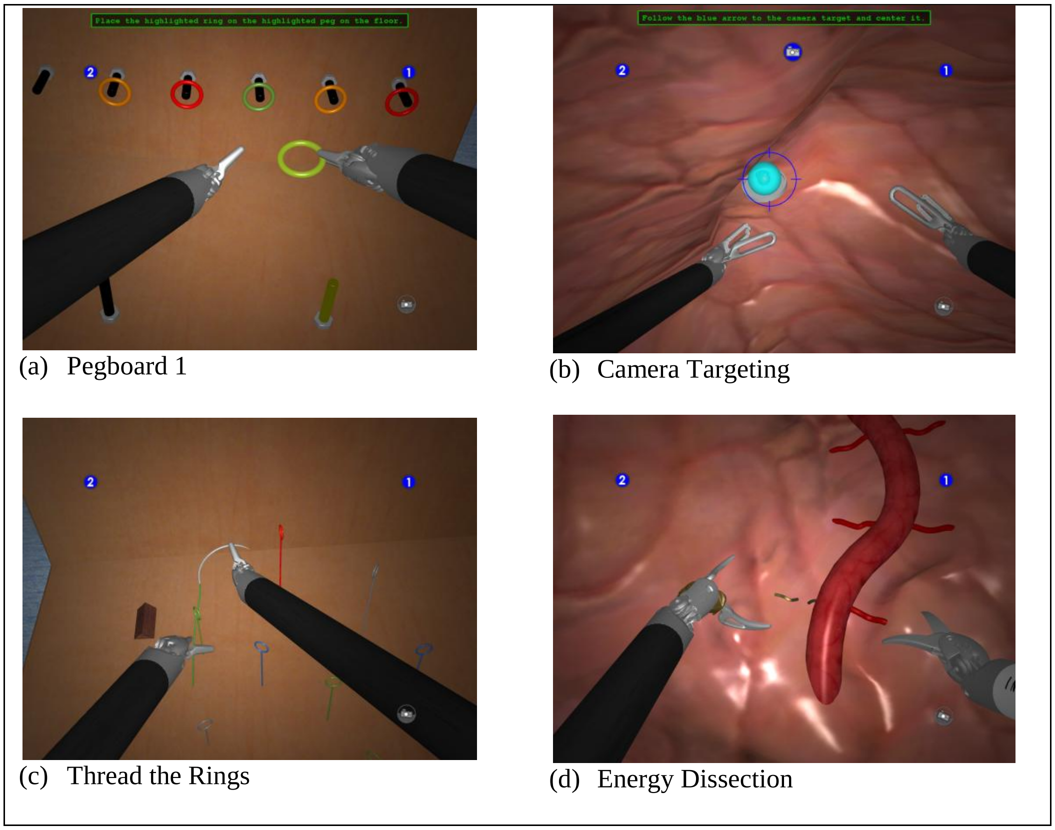


Figure 4. Conceptual Diagram of Communication Latency Thresholds.



(a) Pegboard 1 (b) Camera Targeting

(c) Thread the Rings (d) Energy Dissection

Figure 5. Simulated Surgical Skills Tasks

We further hypothesize that more experienced surgeons will be more successful at managing the effects of latency and would be the best practitioners for this extension of robotic surgery. If this hypothesis is correct, then surgeons with more experience should achieve higher scores and shorter completion times in the simulation experiment that we are performing. This paper reports on the analysis of this specific question comparing surgeon experience to the ability to successfully manage the effects of latency.

In this experiment, subjects performed the four simulated surgical skills exercises shown in Figure 5. These represent many of the core skills that are required in robotic surgery. Each subject performed each exercise three times. First, the subject was given an opportunity to perform the task without any imposed latency. This baseline insured that they were able to successfully operate the controls under normal conditions. Second, they were allowed to perform each of the four exercises at their randomly assigned latency level. These repetitions provided the learning necessary to achieve a sustained level of proficiency within a latent environment (Rayman et al 2006). Finally, each subject performed all four exercises at the same randomly assigned latency level and their performance was measured for analysis in the study.

A single, constant latency level between 100 milliseconds (ms) and 1,000ms at increments of 100ms was randomly assigned to each subject (e.g. 100ms, 200ms, 300ms. 400ms, etc.). A proctor was available to instruct subjects in the use of the equipment and to guide them through the curriculum of the protocol. However, this proctor was not allowed to give suggestions on performance of the exercises or to tell the subject the specific level of latency that they were experiencing.

Data Collection

Experimental data was collected by the simulator software and manually via questionnaires. Research proctors administered a Pre-Test questionnaire on the level of surgical experience and related activities of the subject. All personal and performance data was anonymized to insure that the identity of the subject could not be linked to the data that was collected. The proctors also administered a Post-Test questionnaire at the conclusion of each of the skills exercises during the final performance stage. The simulator software automatically collected multiple measures of the

subject's performance. This provided data for all subjects at zero latency, during their familiarization stage with latency, and during the final stage which is the focus of the analysis. This data will allow us to perform multiple analyses of the skills of robotic surgeons both with and without communication latency, which will be published in future papers.

Pre-Test Questionnaire

The Pre-Test questionnaire identified multiple items of demographic, experience, and practice data on the subjects. These included: age, gender, dominant hand, surgical status, years of surgical experience, years of laparoscopic experience, years of robotic experience, number of weekly procedures in laparoscopy and robotics, and experience with laparoscopic and robotic simulators, as well as with video games and musical instruments. Additional questions captured their opinion on the use of simulation in surgical education and certification.

This data was then matched to the data from their performance in the simulator.

Simulator Performance

During the experiment, the simulator itself collected a number of data points on each subject's performance. These included: time to complete, overall score, total hand motion in centimeters, master working space, number of instrument collisions, number of items dropped, excessive instrument force, distance instruments out of view, incorrect use of electrical energy, simulated blood loss, and number of broken blood vessels.

Post-Test Questionnaire

As the subjects completed their final repetition of each of the four skills exercises, the proctor administered a post-test questionnaire which asked the subject for their opinion on the stress induced by the simulation with latency. This included measures of the mental and physical demands of the task, the pace of the task, their opinion on their level of success, the amount of effort expended, the level of mental discouragement experienced, and their perceived complexity of the exercise.

RESULTS

This paper reports on the analysis of data from the first 54 subjects in the study. Of the 54 subjects who began the experiment, several were unable to complete all of the tasks due to the limited amount of time that they could devote to the experiment. Others found the experiment too taxing and elected to terminate their participation before completion. As a result, we collected complete data sets without latency on 42 subjects and complete data with latency on only 21 of those subjects.

This data was analyzed to determine the level of correlation between the subjects' experience and their performance both with and without latency. For the non-latency sample size of 42 and α=0.05, the Pearson Product Moment Correlation (PPMC) value is 0.304. This means that for a correlation coefficient of two variables in this size of sample to be significant, it must be larger than the PPMC value.

Table 1. Correlation Coefficients without Latency

Exercise	Overall Score	Time Complete
Pegboard 1	0.141	-0.110
Camera Targeting	0.201	-0.173
Thread the Rings	0.156	-0.225
Energy Dissection	0.267	-0.217

In an environment without any latency imposed we found a positive correlation between years of robotic experience and overall performance score, as well as a negative correlation between experience and the total time to complete the exercise (Table 1). Both of these indicate that more experience leads to better performance in the simulator. Though this correlation is consistently supportive that surgeons with more experience perform non-latency exercises better than those with less experience, the degree of this correlation is not large enough to be statistically significant for this sample size.

When latency is added, a simple correlation coefficient is not sufficient for analyzing the effect of robotic experience on performance. Each subject received a randomly assigned latency, of which there were 10 possibilities. Within the current sample, we have between 0 and 5 subject data points at each latency level. Therefore, under latency, we examine the data by visual examination of a multiline scatter plot.

Scatterplots can illustrate the linear relationship between two variables in the model. Without latency, a relationship can be seen for both overall performance score and time to complete the exercise (Figures 6 & 7). However, when latency is present, the plots show that there is not a relationship between the two variables for the subjects tested (Figures 8 & 9).

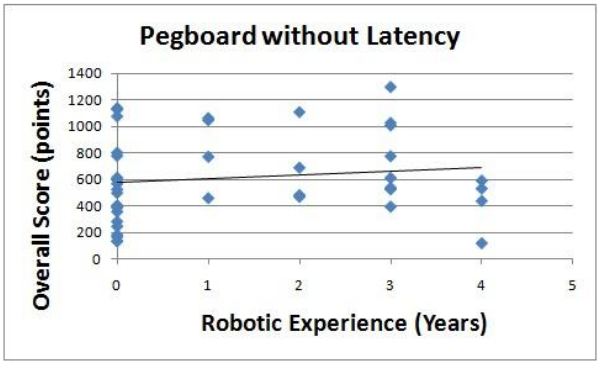


Figure 6. Correlation between Robotic Experience and Overall Score for the Peg Board exercise without communication latency.

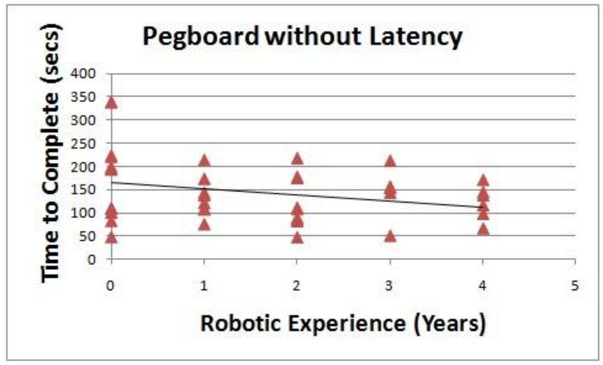


Figure 7. Correlation between Robotic Experience and Time to Complete for the Peg Board exercise without communication latency.

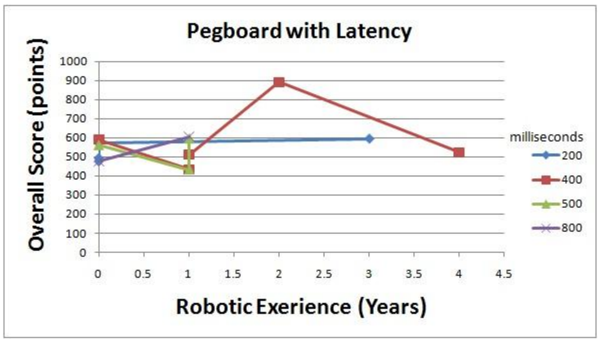


Figure 8. Correlation between Robotic Experience and Overall Score for the Peg Board exercise with various communication latencies.

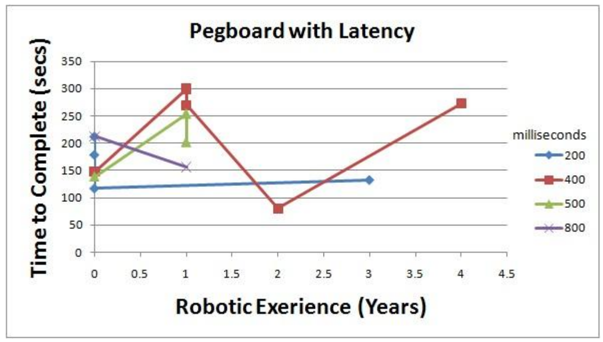


Figure 9. Correlation between Robotic Experience and Time to Complete for the Peg Board exercise with various communication latencies.

The data suggests that surgeons who have more experience in robotic surgery are not better equipped to self-manage the challenges presented by communication latency in telesurgery. Subjects with little experience are as likely to successfully manage latency as are surgeons with more experience.

This same trend holds when comparing independent variables like total surgical experience and laparoscopic experience to the scores achieved in the simulator with latency.

CONCLUSIONS

The lack of correlation between experience and telesurgical performance under latency refutes our original hypothesis that a more experienced surgeon would more successfully manage the effects of latency. This negative finding has led to speculation on the cause of these results. Several may be possible, but each will require additional experimentation. First, experienced surgeons may be very talented, but fixed, in their methods of performing surgery. This may lead them to perform poorly under latency because it is difficult for them to modify their behaviors, where inexperienced surgeons are less ingrained and more adaptable to the situation. Second, since the simulator is a computer-generated virtual environment, it is possible that surgeons who have more experience in simulators, virtual worlds, and computer games may have developed a proficiency for solving problems in this kind of environment. They may also have experienced latency in those environments and developed techniques for compensating for it. Third, the ability to manage latency may be related to the physical and biological wiring of an individual. This could be a similar phenomenon to the tendency for some people to experience simulator sickness, while others do not suffer from it. These speculations are worthy of further investigation.

The objective of this analysis was to identify the degree to which a surgeon can compensate for the effects of latency that are present in a telesurgery environment. The long-term goal is to identify the thresholds where safe and successful surgery can be performed. Our findings at this point refute our hypothesis that more experienced surgeons would be able to manage latency more successfully. In the data collected there is no correlation between robotic experience and the ability

to achieve a higher score in the simulator when latency is inserted into the procedure.

These results may inform research on remote teleoperation in other environments, such as the control of UAV's and UGV's. Experienced pilots and vehicle drivers may not be better equipped to manage the effects of latency than pilots/drivers with less experience. Other factors may be more important in predicting a person's ability to tele-operate a remote system successfully. The similarity between remote surgery and remote vehicle operation is speculative and would require specific research experiments with those systems to verify.

ACKNOWLEDGEMENTS

This effort was sponsored by the Department of the Army, Award Number W81XWH-11-2-0158 to the recipient Adventist Health System/Sunbelt, Inc., Florida Hospital Nicholson Center. "The U.S. Army Medical Research Acquisition Activity, 820 Chandler Street, Fort Detrick MD 21702-5014 is the awarding and administering acquisition office." The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

REFERENCES

Anvari M, Broderick T, Stein H, Chapman T, Ghodoussi M, Birch DW, McKinley C, Trudeau P, Dutta S, & Goldsmith P. (2005) Impact of latency on surgical precision and task completion during robot assisted remote telepresence surgery. *Computer Aided Surgery*, 10(2):93-9.

Anvari M. (2007) Remote telepresence surgery: A Canadian experience. *Surgical Endoscopy*, 21:537-41

Fabrizio MD, Lee BR, Chan DY, Stoianovici D, Jarrett TW, Yang C, & Kavoussi L. (2000). Effect of time delay on surgical performance during telesurgical performance and during telesurgical manipulation. *Journal of Endourology*, 14:133-8

FDA. Computer-Assisted Surgery: An Update. (2005). Available online at: http://www.fda.gov/fdac/features/2005/405_computer.html

Himpens J, Leman G, & Cardiere GB. (1998). Telesurgical laparoscopic cholecystectomy. *Surgical Endoscopy*, 12:1091

Hung, A, Zehnder P, Patil M, Cai J, Ng A, Aron M, Gill I, & Desai M. (2011). Face, Content and Construct Validity of a Novel Robotic Surgery Simulator, *Journal of Urology*.

Janetschek G, Bartsch G, & Kavoussi LR. (1998). Transcontinental interactive laparoscopic telesurgery between the United States and Europe. *Journal of Urology*, 160:1413.

Kenney P, Wszolek M, Gould J, Libertino J, & Moinzadeh A. Face. (2009). Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery, *Journal of Urology*, 73(6).

Kwoh, Y. S., Hou, J., Jonckheere, E. A. & Hayall, S. (1988). A robot with improved absolute positioning accuracy for CT guided stereotactic brain surgery. *IEEE Transactions on Biomedical Engineering*, 35(2), 153–161.

Lum M.J.H. et al (2009) Teleoperation in Surgical Robotics - Network Latency Effects on Surgical Performance. *Proceedings of the 31st Annual International Conference of the IEEE EMBS*, Minneapolis, Minnesota, USA.

Marescaux J, Leroy J, Rubino F, Smith M, Vix M, Simone M, Mutter D. (2002). Transcontinental robot-assisted remote telesurgery: feasibility and potential applications. *Annals of Surg*ery, 235:487-492.

Rayman R, Croome K, Galbraith N, McClure R, Morady R, Peterson S, Smith S, Subotic V, Van Wynsberghe A, Primak S. (2006). Long distance robotic telesurgery: A feasibility study for care in remote environments. *International Journal of Medical Robotics*, 2(3):216-24.

Satava R & Green P. (1992). The next generation Telepresence Surgery: Current status and implications for endoscopy. *Gastrointestinal Endoscopy*, 38:277.

Sterbis, J.R. et al. (2007). Transcontinental telesurgical nephrectomy using the da Vinci robot in a porcine model. *Journal of Urology*, 11(27), 971-973.

Robotic Surgical Education with Virtual Simulators

Roger Smith, PhD
Florida Hospital Nicholson Center
Celebration, FL
roger.smith@flhosp.org

Mireille Truong, MD
Florida Hospital Celebration Health
Celebration, FL
mireille.truong@gmail.com

ABSTRACT

The rapid advancement of robotic surgical technology and its implementation in minimally invasive surgical procedures has led to the need to develop more efficient and effective training methods, as well as assessment and skill maintenance tools for surgical education. Previous studies have shown that virtual simulation training is effective for improving laparoscopic surgical performance. However, few have evaluated the effectiveness of these types of simulators for improving robotic surgery proficiency.

A three-part evaluation of the available robotic simulators is being performed to address the value and possible applications of the devices. The first part is an objective review and comparison of the design and capabilities of all of the simulators, which provides base specifications to aid potential users with selection of the device that best meets their needs. The second part is a subjective opinion on the usability of the simulators, which will include a survey of various health professionals and medical students without prior experience using the simulation devices. The third part includes a two-month experiment to determine which simulator has the greatest positive impact on robotic surgical performance and the degree of skill retention over a period of inactivity.

This paper describes the results of the first part of this study. It provides comparative data on all three simulators - the da Vinci Skills Simulator (Intuitive Surgical Inc.); dV-Trainer (Mimic Technologies, Inc.); and RoSS (Simulated Surgical Skills LLC). This includes details about the curriculum, scoring method, system administration, visual resolution, validation, and support tools for the devices.

ABOUT THE AUTHORS

Roger Smith, PhD, is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading technology implementation. This includes identifying, executing and managing industry, military and federally funded simulation, modeling and training projects. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *Innovation for Innovators: Leadership in a Changing World.* He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

Mireille Truong, MD, is a Minimally Invasive Gynecology Fellow at Celebration Health Florida Hospital, a Research Physician at the Nicholson Center and an Assistant Professor of Obstetrics and Gynecology (OB/GYN) at the University of Central Florida's College of Medicine. During her residency in OB/GYN at the University of Illinois at Chicago, she received a number of awards and honors, including Administrative Chief Resident, the American Association of Gynecologic Laparoscopists' Special Resident in Minimally Invasive Gynecology Award and Best Overall Excellence in Gynecologic Care Award. She has dedicated her time to education of medical students and physicians via an array of academic and teaching appointments, professional organization board positions, peer-reviewed publications and presentations at various national and international conferences. Her research interests include surgical education, robotic simulation and minimally invasive gynecologic surgery.

Robotic Surgical Education with Virtual Simulators

Roger Smith, PhD
Florida Hospital Nicholson Center
Celebration, FL
roger.smith@flhosp.org

Mireille Truong, MD
Florida Hospital Celebration Health
Celebration, FL
mireille.truong@gmail.com

BACKGROUND

For every complex and expensive system there emerges a need for training devices and scenarios that will assist new learners in mastering the use of the device and understanding how to apply it with value. In laparoscopic surgery, simulators have played an important role in improving the practice of surgery over the last 20 years (Schout, 2010; Wohaibi, 2010 et al). The same trends and values will likely apply to robotic surgery with the increased use of robotic technology for a growing variety of minimally invasive surgical procedures. The complexity, criticality, and cost associated with the effective application of the da Vinci surgical robot have stimulated the commercial creation of simulators which replicate the operations of this robot. The objective of this paper is to evaluate and compare the three commercially available robotic simulators shown in Figure 1:

- da Vinci Skills Simulator (Intuitive Surgical Inc.);
- dV-Trainer (Mimic Technologies, Inc.); and
- RoSS (Simulated Surgical Skills LLC).

Each of these possesses unique traits which make them valuable solutions for different types of users and learning environments.

DVSS dV-Trainer RoSS

Figure 1. Simulators of the da Vinci surgical robot

METHODS

Florida Hospital Nicholson Center owns and uses all three of these simulators. This cross-device access and experience is rare and provides unique comparative insight into the capabilities of all of the devices. We reviewed the users' manuals for the devices to collect details about each system and performed our own experiments with each device to create comparative materials across all devices.

We performed a systematic literature review on all three devices. The PubMed database of medical research was searched for all references to the devices through February 2013. References from retrieved articles were reviewed to broaden the search. The data extracted from these studies include training exercise modules, scoring systems, costs, educational impact and validation methods. We identified 32 studies investigating simulation in robotic surgery.

Finally, we submitted our comparative data on the systems to the manufacturers of each device to receive a review of the accuracy of the information.

The result of this work in this comparative review of the devices which evaluates the characteristics, exercise modules, scoring systems, costs, validity, advantages and disadvantages of each simulator.

RESULTS

Each of these devices is manufactured by a different company and provides a unique hardware and software solution for training and surgical rehearsal. The capabilities and features of each are summarized in Table 1.

Capabilities and Features

Da Vinci Skills Simulator (Intuitive Surgical Inc.)

The da Vinci Skills Simulator (DVSS) consists of a customized computer package that attaches to the back of the surgeon’s console of an actual da Vinci Si robot. This simulator connects to the surgeon’s console via a single proprietary networking cable identical to that used to connect the components of the actual robotic surgical system.

Advantages

Attached simulators of this type are usually referred to as “embedded trainers” because they take advantage of the equipment that has already been constructed, purchased, and installed for the operation of the real system. These kinds of simulators are especially common in military facilities which face limited space and weight constraints. They can significantly reduce the hardware that must be purchased solely for simulation purposes. The U.S. Navy uses these kinds of simulators aboard ships to reduce weight and space requirements, enabling them to train while the ship is at sea.

Another significant advantage of an attached simulator is that it allows the trainee to use the actual controls from the real system to control the simulation. This insures that the training experience is almost identical in feel to the real system, which can contribute to higher transfer of skills from the training sessions to the real system. Additionally, this minimizes the amount of time spent learning the unique functionalities of the simulator device and allows the trainee to focus the majority of his/her learning experience on skills acquisition and attaining proficiency. Finally, there is the cost advantage for the simulator device itself. Because much of the hardware and software expenses are already embedded in the real system, the simulator can be very economical to purchase.

Table 1. Robotic Simulator Feature Comparison

Features	DVSS	dV-Trainer	RoSS
System Manufacturer	Intuitive Surgical Inc.	Mimic Technologies Inc.	Simulated Surgical Systems LLC
Specifications (Simulator only)	Depth 7” Height 25” Width 23” 120 or 240V power	Depth 36” Height 26” Width 44” 120 or 240V power	Depth 44” Height 77” Width 45” 120 or 240V power
Specifications (Complete System as shown in Figure 1)	Depth 41” Height 65” Width 40” 120 or 240V power	Depth 36” Height 59” Width 54” 120 or 240V power	Depth 44” Height 77” Width 45” 120 or 240V power
Visual Resolution	VGA 640 x 480	VGA 640 x 480	VGA 640 x 480
Components	Customized computer attached to da Vinci surgical console	Standard computer, visual system with hand controls, foot pedals.	Single integrated custom simulation device
Support Equipment	da Vinci surgical console, custom data cable	Adjustable table, touch screen monitor, keyboard, mouse, protective cover, custom shipping container	USB adapter, keyboard, mouse
Exercises	35 simulation exercises	51 simulation exercises	52 simulation exercises.
Optional Software	PC-based Simulation management	Mshare curriculum sharing web site	Video and Haptics-based Procedure Exercises (HoST)
Scoring Method	Scaled 0-100% with passing thresholds in multiple skill areas	Proficiency-based point system with passing thresholds in multiple skill areas	Point system with passing thresholds in multiple skill areas
Student Data Management	Custom control application for external PC. Export via USB memory stick.	Export student data to delimited data file.	Export student data to delimited data file.
Curriculum Customization	None	Select any combination of exercises. Set passing thresholds and conditions.	Select specifically grouped exercises. Set passing thresholds.
Administrator Functions	Create student accounts on external PC. Import via USB memory stick.	Create student accounts. Customize curriculum.	Create student accounts. Customize curriculum.
System Setup	None.	Calibrate controls.	Calibrate controls.
System Security	Student account ID and password.	PC password, Administrator password, Student account ID and password.	PC password, Administrator password, Student account ID and password.
Simulator Base Price	$85,000	$95,000	$107,000
Support Equipment Price	$502,000	$9,100	$0
Total Functional Price	$587,000	$104,100	$107,000

Disadvantages

Attached simulators like the DVSS also come with inherent disadvantages to balance their positive traits.

The largest drawback is the availability and accessibility of a simulator which requires the real robotic system. An attached DVSS simulator cannot be used without access to a real surgeon’s console and therefore is only available for use when the robotic system is not in use. This implies that the trainee would only be able to use the simulator outside of normal operating room working hours and would need logistical access to the robot and the simulator. da Vinci robots are expensive devices which hospitals typically attempt to maximize use of in order to recoup their investment. In a very active surgical hospital, it can be difficult to obtain access to a surgeon’s console to support training with this simulator.

The DVSS is designed to connect to the surgeon’s console using the same proprietary networking cable that connects the major robot components. This makes the attachment and set-up process very easy for clinicians to master. However, it also means that the DVSS can only be used with the Si model surgeon’s console. The previous S and Standard models use a different set of cables, which are not compatible with the simulator.

Similar to the military’s experience with embedded and attached simulators, heavy usage of the DVSS comes with a corresponding heavy use of the surgeon’s console. The Army and Navy have discovered that these types of

simulators put more usage hours on real equipment controls which lead to more maintenance costs for those devices. Given the possibility of regular and continuous simulation training with such as device in addition to actual surgical usage, the real equipment experiences usage rates that can be many times higher than normal for the equipment. Since the da Vinci systems operate under a maintenance contract that covers all services, the additional costs of maintenance are not born by the hospital owner, but by the equipment vendor. The primary impact to the owner would only be in the area of availability for both real surgeries and training events due to downtime associated with maintenance.

As mentioned under advantages, the cost of an attached simulator is typically much lower than other forms. However, this is countered by the fact that the customer must purchase or have available a real piece of equipment to support the use of the simulation.

dV-Trainer (Mimic Technologies Inc.)

The dV-Trainer is a separate, stand-alone simulator of the da Vinci robot. The surgeon's console, controls, and vision cart are mimicked in hardware, while a 3D software model replicates the functions of the robotic arms and the surgical space.

Mimic also developed the core simulator software for the DVSS and used the same package in version 1.0 of their own dV-Trainer. As a result, the exercises in those versions of the systems are nearly identical. The current version 2.0 of the dV-Trainer has a number of new exercises, which are not found in the DVSS, and the graphics have been upgraded so the visual presentation is no longer identical. The differences in visual presentation can be seen in the figures later in the paper.

The dV-Trainer consists of three major pieces of equipment and a number of smaller support pieces. The largest pieces are the "Phantom" hood which replicates the vision and hand controls of the da Vinci surgeon's console, the foot pedals of the surgeon's console, and a high-performance desktop computer which generates the 3D images and calculates the interactions with the surgeon's controls. Smaller support equipment includes a touch screen monitor, keyboard, and mouse to enable an instructor to guide the student through exercises and allow an administrator to manage the data that is collected.

Because the dV-Trainer replicates both the hardware and software of the da Vinci robot, it is a much larger system than the DVSS alone, though smaller than a real surgeon's console with the DVSS attached. It has the advantage of providing a training system that is completely independent of the need for any piece of the real surgical robot. The simulator can be configured to imitate either the S or the Si model of the da Vinci robot.

The disadvantage of this kind of system is that the simulated hardware is somewhat different than the real equipment and does not exactly replicate the feel of the real physical equipment. There is always a trade-off between lower price and perfect accuracy of a simulator. Also, the simulator must be updated separately when the real equipment is modified.

Robotic Surgical System (Simulated Surgical Systems LLC)

The RoSS is also a complete, stand-alone simulator of the da Vinci robot. This device is designed as a single piece of hardware that has a similar design to the surgeon's console of the robot. The hardware device includes a single 3D computer monitor, hand controls that are modified commercial force feedback devices, pedals that replicate either the S or the Si model of the da Vinci robot, and an external monitor for the instructor. Customers must purchase either the S or Si version of the device.

The company has developed a set of 3D virtual exercises that are unique from those found in both of the other simulators. They also provide an optional video-based surgical exercise in which the user is guided through the movements necessary to complete an actual surgical procedure. At this writing, these modules are available for radical prostatectomy, cystectomy, and hysterectomy. These guided videos take advantage of the force feedback capabilities of the hand controllers to push and pull the student's hands to follow the simulated instruments on the

screen. They require the student to perform specific movements accurately during the video before the operation will proceed.

Exercise Modules

The exercise modules in each simulator are organized into hierarchical menus according to the surgical skill being addressed and the complexity of the exercise (Table 2). Each simulator provides on-system instructions for each exercise in the form of textual documents and narrated step-by-step video demonstrations. Upon completion of each exercise, the system automatically proceeds to a scoreboard showing the student's performance on the exercise.

Table 2. Comparative Simulator Exercise Categories

DVSS	dV-Trainer	RoSS
Surgeon Console Overview	Surgeon Console Overview	Orientation Module
Endowrist Manipulation	Endowrist Manipulation 1	Motor Skills
Camera and Clutching	Endowrist Manipulation 2	Basic Surgical Skills
Energy and Dissection	Camera and Clutching	Intermediate Surgical Skills
Needle Control	Energy and Dissection	Hands-on Surgical Training
Needle Driving	Needle Control	
Troubleshooting	Needle Driving	
Games	Games	
Suturing Skills	Suturing Skills	

DVSS

The DVSS contains 35 exercises organized into nine categories. These begin with introductory video and audio instructions on how to use the robotic equipment, and move through progressively more difficult skills.

dV-Trainer

Most of the simulation software for Intuitive's DVSS was developed by Mimic Technologies. Therefore, version 1.0 of the DVSS and the dV-Trainer contained nearly identical exercises, closely matching menu systems, and identical scoring mechanisms. However, over time the two sets of software have diverged and the current versions of the simulators differ in functionality and appearance. The current version of the dV-Trainer (v 2.0) contains 51 exercises organized into nine categories.

Though many of the exercises are identical between the DVSS and the dV-Trainer, the graphics resolution and details have been improved in version 2.0 of the dV-Trainer software. Since this system is driven by a commercial PC which can be upgraded rather easily, it is possible for the software to evolve and be replaced more easily than for a custom hardware package like the DVSS which would require upgrades to some of the components inside the device.

RoSS

The RoSS simulator contains 52 unique exercises, organized into 5 categories, and arranged from introductory to more advanced, just as in the other two simulators. The RoSS system of exercises is unique in that they list fewer named exercises, but provide three different difficulty levels for most of them (i.e. Level 1 is the easiest, Level 2 is intermediate, and Level 3 is advanced).

The RoSS contains a unique capability that is not found in either of the other simulators called "Hands-on Surgical Training" or "HoST." This is an integration of surgical skills exercises with a video of an actual surgery. Videos of actual surgical procedures play in the surgeon's visual space, overlaid with animated icons which instruct the student to perform specific actions during the progression of the surgery video. The necessary actions are prompted with audio instructions. For the HoST exercise to progress, the student must perform the specific actions at specific times. The simulator will pause the video and allow the student to repeat the action until it is performed as required by the instructions.

The hand controllers of the RoSS simulator are modified versions of a commercially available 3D haptic input device called the Omni Phantom™. This product uses internal motors and gears to apply haptic feedback to the hand movements of the user. For the HoST exercises, the simulator uses this capability to move the student's hands in sync with the movements of the surgeon's instruments in the master video.

Proficiency Scoring System

Each of the three simulators provides a different scoring method. All three use the host computer to collect data on the performance of the student at the controls in multiple performance areas. With this data, they provide a score for specific performance traits, as well as combining all of these into a single composite score of performance for the entire exercise. The algorithm used to create this composite score is described in the user's manuals of each of the simulators. Examples of each of these scoreboards are shown in Figure 2.

DVSS	dV-Trainer	RoSS
		

Figure 2. Example Scoreboards from Each Simulator

In addition to the objective metrics that can be collected by the computer, the developers of each simulator have been challenged to provide thresholds which indicate whether the student's score is considered a "passing" or "failing" performance. All three have identified threshold scores which would indicate acceptable and warning scoring levels. These are commonly interpreted as "passing" (above acceptable threshold) and "failing" (below warning threshold), with a "warning" area between the two thresholds. These thresholds create green, yellow, and red performance areas, which can be used to visually communicate the quality of the student's performance in each area of measurement. Each simulator also provides a single composite score for the entire exercise.

DVSS

The DVSS performance scoring method has a number of metrics, which are applied to every exercise and others which are only used for exercises in which they are relevant. Table 3 presents the metrics, which are applicable to all exercises. For details on the more specialized metrics, the reader may consult the user's manual for the simulator.

Because the DVSS is a closed, turn-key system with an ease of use similar to the actual surgical robot, most of the data displays and threshold adjustments found in the other simulators are not available in this device. Most simulator settings are determined by the manufacturer and cannot be changed by the user.

Table 3. DVSS and dV-Trainer Scoring Method

Overall Score	Composite evaluation of the exercise performance.
Time to Complete	Number of seconds to complete the exercise.
Economy of Motion	Number of centimeters of instrument tip movement.
Instrument Collisions	Number of times that the instruments touched each other.
Excessive Instrument Force	Number of seconds that excessive robotic force was applied against objects in the environment.
Instrument Out of View	Number of centimeters that an instrument tip moved outside of the viewing area.
Master Workspace Range	Radius in centimeters than contains the movement of the instrument tips.
Drops	Number of objects dropped from the grasp of the instruments.

dV-Trainer

Originally, the DVSS and the dV-Trainer shared the same scoring method, but more recent versions of the dV-Trainer offer both this original “version 1.0” scoring method, as well as a new “version 2.0” method based on the proficiency measured from experienced surgeons. The skills measured are the same (Table 2), but the interpretation of those into a score is different. The instructor can select the preferred scoring method for each curriculum that is constructed in the dV-Trainer.

Users will notice that the newer scoring method uses total points earned rather than percentages. The passing and warning thresholds can be adjusted by the administrator. The philosophy, validity, and effects associated with these settings are more detailed than is necessary for understanding the use of the simulator. Interested readers should consult the user’s manual and published literature for details on the two scoring mechanisms.

RoSS

The principles behind the scoring system on the RoSS are the same as those for the DVSS and the dV-Trainer. However, most of the metrics collected are different. The standard measurements are shown in Table 4.

Table 4. RoSS Scoring Method

Overall Score	Composite evaluation of the exercise performance.
Camera Usage	Optimal movement of camera.
Left Tool Grasp	Optimal number of tool grasps with left hand tool.
Left Tool Out of View	Distance left hand tool is out of view
Number of Errors	Number of collision or drop errors in an exercise.
Right Tool Grasp	Optimal number of tool grasps with right hand tool.
Right Tool Out of View	Distance right hand tool is out of view.
Time	Time to complete the exercise.
Tissue Damage	Number of times that instruments damaged tissue with excessive force or unnecessary touches.
Tool-Tool Collision	Number of times tools touched each other.

Like each of the other simulators, there are multiple displays of the performance data for a student. The initial display presented at the completion of an exercise shows a horizontal bar which is colored green, yellow, or red to indicate passing or failing. The magnitude of the bar is a rough measure of the quality of performance. Additional displays show the numeric score and its relative position to a passing threshold.

Validation of Devices

Validation studies serve to determine whether a simulator can actually teach or assess what it is intended to teach or assess. In medical simulation, there are generally accepted validity classifications, which include face, content, construct, concurrent and predictive validity (McDougall, 2007). Face and content validity are considered subjective approaches while the other three are objective approaches to validation.

Table 5 provides a summary of the published validation studies for these simulators. All three have publications establishing face, content, construct, and concurrent validation. There is only one published study on the predictive validity of the DVSS (Hung, 2012). Recent presentations also explore the validity of the RoSS curriculum (Stegemann, 2013) and the RoSS’ HoST procedural modules (Ahmed, 2013).

Table 5. Validation of robotic surgical simulators

Validation	DVSS	dV-Trainer	RoSS
Face	*Hung 2011* *Kelly 2012* *Liss 2012*	*Lendvay 2008* *Kenney 2009* *Sethi 2009* *Perrenot 2011* *Korets, 2011* *Lee 2012*	*Seixas-Mikelus 2010* *Stegemann, 2012*
Content	*Hung 2011* *Kelly 2012* *Liss 2012*	*Kenney 2009* *Sethi 2009* *Perrenot 2011* *Lee 2012*	*Seixas-Mikelus 2010* *Colaco, 2012*
Construct	*Hung 2011* *Kelly 2012* *Liss 2012* *Finnegan 2012*	*Kenney 2009* *Korets, 2011* *Perrenot 2011* *Lee 2012*	*Raza, 2013*
Concurrent	*Hung 2012*	*Lerner 2010* *Perrenot 2011* *Korets 2011* *Lee 2012*	*Chowriappa, 2013*
Predictive	*Hung 2012*		

CONCLUSIONS

The three simulators described in this review article are complex systems, which are significantly less costly than the actual da Vinci robotic surgical system and can be operated at a fraction of the cost of the instruments required for this robot. There are currently no available studies comparing the three simulators head-to-head and therefore until those studies are performed, no universal recommendation can be made for one device over the other, but rather a decision to use one simulator over the other should be based on unique and individual needs.

This article represents the first part of a comprehensive analysis of robotic surgical simulators. The second part is a subjective opinion survey on the usability of the simulators. Subjects for this survey will include attending surgeons, fellows, residents, and medical students without prior experience using the simulation devices. The third part will include a select group of surgical fellows will participate in a two-month experiment practicing on one of the simulators while their performance is measured every two weeks to assess for changes and maintenance of skill levels. The experiment is designed to determine which simulator has the greatest positive impact on robotic surgical performance and the degree to which those improvements are retained across a period of inactivity.

ACKNOWLEDGEMENTS

The authors want to thank the following representatives of each simulator company for their assistance in collecting data, images, and videos, as well as verifying the accuracy of the contents of the paper. Peter Dominick, Intuitive Surgical Inc.; Jeff Berkley, Mimic Technologies, Inc.; and Kesh Kesavadas, Simulated Surgical Systems, LLC.

REFERENCES

Ahmed K, Chowriappa A, Din R, Field E, Raza SJ, Fazili A, Samarasekera D, Kaouk J, Eichel L, Joseph J, Ghazi A, Mohler J, Peabody J, Eun D, Shi Y, Wilding G, Kesavadas T, Guru KA. (2013). A Multi-institutional Randomized Controlled Trial of an Augmented-Reality based technical skill acquisition module for Robot-assisted Urethro-vesical Anastomosis. *Podium presentation, AUA Annual Congress, San Diego.*

Berry E. (2010). Robotic Surgical Simulator advances training criteria for today's surgeons. *Briefings on Credentialing,* 19(5).

Burden C, Oestergaard J, Larsen C. (2011) Integration of laparoscopic virtual-reality simulation into gynaecology training. *British Journal of Obstetrics and Gynecology,* 118(3): 5–10.

Chowriappa A, Raza SJ, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody JO, Menon M, Kesavadas T, Guru KA. (2013). Development & Validation of a Composite Grading & Scoring System for Robot-assisted Surgical Training – The Robotic Skills Assessment Scale. *Podium presentation, AUA Annual Congress, San Diego*.

Colaco M, Balica A, Su D. (2012) Initial experiences with RoSS surgical simulator in residency training: a validity and model analysis, *Journal of Robotic Surgery*, 7(1):71-75

dV-Trainer Robotic Simulator Users Manual, Mimic Technologies, Inc. (2012).

Finnegan KT, Meraney AM, Staff I. (2012). da Vinci Skills Simulator construct validation study: correlation of prior robotic experience with overall score and time score simulator performance. *Journal of Urology,* 80(2):330-5.

Guru K, Baheti A, Kesavadas T, Kumar A, Srimathveeravalli G, and Butt Z. (2009). In-Vivo Videos Enhance Cognitive Skills for da Vinci® Surgical System. *Journal of Urology* 181:4 Supplement.

Hart R, Karthigasu K. (2007). The benefits of virtual reality simulator training for laparoscopic surgery. *Current Opinion in Obstetrics and Gynecology*, 19:297– 302.

Hung AJ, Patil MB, Zehnder P. (2012). Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. *Journal of Urology,* 187(2):630-7.

Hung AJ, Zehnder P, Patil MB. (2011). Face, content and construct validity of a novel robotic surgery simulator. *Journal of Urology,*186(3):1019-24.

Kelly DC, Margules AC, Kundavaram CR. (2012). Face, content, and construct validation of the da Vinci Skills Simulator. *Journal of Urology,* 79(5):1068-72.

Kenney P, Wszolek M, Gould J, Libertino J, Moinzadeh A. (2009). Face, Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery. *Journal of Urology*, 73(6):1288-92.

Kesavadas T , Kumar A, Srimathveeravalli G, Karimpuzha S, Baheti A, Chandrasekhar R, Wilding G, Butt Z and Guru K. (2009). Efficacy of Robotic Surgery Simulator (RoSS) for the daVinci® Surgical System. *Journal of Urology* 181:4 Supplement.

Kesavadas T, Stegemann A, Sathyaseelan, G, Chowriappa A, Srimathveeravalli G, Seixas-Mikelus S, Chandrasekhar R, Wilding G, and Guru K (2011). Validation of Robotic Surgery Simulator (RoSS). *Journal of Student Health Technology Informatics*, 163:274-6.

Kim S, Cho J. (2011). Validation Study of a 3D Virtual Robot Simulator as a Robot Surgery Training System. *Journal of Korean Colorectal Surgery*.

Korets R, Graversen JA, Mues A, Gupta M, Landman J, Badani KK. (2011). Face and construct validity assessment of 2nd generation robotic surgery simulator. *Journal of Urology,* 185 (Suppl.): e488.

Korets R, Mues AC, Graversen JA, Gupta M, Benson MC, Cooper KL, Landman J, Badani KK (2011) Validating the Use of the Mimic dV-Trainer for Robotic Surgery Skill Acquisition Among Urology Residents. *Journal of Urology*, 78(6):1326-30.

Kumar A, Seshadri S, Baheti A, Srimathveeravalli G, Butt Z, Kuvshinoff B, Mohler J, Kesavadas T and Guru K. (2008). Virtual Reality Surgical Trainer for Robotic Urological Procedures. *Journal of Urology* 179:4:Supplement 1.

Lee JY, Mucksavage P, Kerbl DC. (2012) Validation study of a virtual reality robotic simulator--role as an assessment tool? *Journal of Urology*, 187(3):998-1002.

Lendvay TS, Casale P, Sweet R. (2008). VR robotic surgery: randomized blinded study of the dV-Trainer robotic simulator. *Journal of Student Health Technology Informatics*, 132:242-4.

Lendvay, P. Casale, R. Sweet, C. Peters. (2008). Initial validation of a virtual-reality robotic simulator**.** *Journal of Robotic Surgery*, 2(3):145-149.

Lerner M, Ayalew M, Peine W, Sundaram C. (2010). Does Training on a Virtual Reality Robotic Simulator Improve Performance on the da Vinci® Surgical System?. *Journal of Endourology*, 24(3):467-72.

Liss MA, Abdelshehid C, Quach S. (2012). Validation, Correlation, and Comparison of the da Vinci Trainer(™) and the da Vinci Surgical Skills Simulator(™) Using the Mimic(™) Software for Urologic Robotic Surgical Education. *Journal of Endourology*, 26(12):1629-34.

McDougall EM. (2007). Validation of surgical simulators. *Journal of Endourology*, 21(3):244-7.

Munro MG. (2012). Surgical simulation: where have we come from? Where are we now? Where are we going?. *Journal of Minimally Invasive Gynecology*, 19(3):272-83.

Perrenot C, Perez M, Tran N. (2012) The virtual reality simulator dV-Trainer(®) is a valid assessment tool for robotic surgical skills. Journal of Surgical Endoscopy, 26(9):2587-93.

Raza SJ, Chowriappa A, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody J, Menon M, Kesavadas T, Guru KA. (2013). Construct Validation of the Fundamental Skills of Robotic Surgery (FSRS) Curriculum. *Podium presentation, AUA Annual Congress, San Diego,*

Robotic Surgery Simulator Users Manual: For Models S and Si, Simulated Surgical Systems LLC, 2012.

Schout BM, Hendriks AJ, Scheele F et al. (2010). Validation and implementation of surgical simulators: a critical review of present, past, and future. *Journal of Surgical Endoscopy*, 24(3):536-46.

Seixas-Mikelus S, Adal A, Srimathveeravalli G, Kesavadas T, Baheti A, Chandrashekar R, Wilding G and Guru K. (2010). Can Image Based Virtual Reality Help Teach Anatomy? *Journal of Endourology*, 24(4):629-34.

Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G. (2010). Face validation of a novel robotic surgical simulator. *Journal of Urology*, 76(2):357-60.

Seixas-Mikelus SA, Stegemann AP, kesavadas T. (2011). Content validation of a novel robotic surgical simulator. *Journal of the British Association of Urological Surgeons*, 107(7):1130-5.

Sethi AS, Peine WJ, Mohammadi Y. (2009). Validation of a novel virtual reality robotic simulator. *Journal of Endourology*, 23(3):503-8.

Sethi AS, Peine WJ, Mohammadi Y. (2009). Validation of a novel virtual reality robotic simulator. *Journal of Endourology*, 23(3):503-8.

Skills Simulator for the da Vinci Si Surgical System, (Users Manual), Intuitive Surgical Inc., 2012.

Smith, M. L. (2011). Simulation and education in gynecologic surgery. *Journal of Obstetrics and Gynecology Clinics of North America*, 38:733-740.

Stegemann A, Kesavadas T, Rehman S, Sharif M, Rao M, duPont N, Shi Y, Wilding G, Hassett J, and Guru K. (2013). Development, Implementation, and Validation of a Simulation-Based Curriculum for Robot-Assisted Surgery. *Journal of Urology*, 81(4):767-74.

Stegemann AP, Kesavadas T, Rehman S, Sharif M, Rao A, DuPont N, Shi Y, Wilding,G, Hassett J, and Guru K. (2012). Development, Implementation and Validation of a Simulation-Based Curriculum for Robot-Assisted Surgery. *Presented at the AUA poster session, May 19-23, 2012 Atlanta, GA.*

Su D, and Barone J. (2011). Initial Experience with the ROSS Robotic Simulator in Residency Training. *Moderated poster, AUA 2011.*

Wohaibi EM, Bush RW, Earle DB et al. (2010). Surgical resident performance on a virtual reality simulator correlates with operating room performance. *Journal of Surgical Research* 160(1):67-72.

Zorn K, and Gautam G. (2011). Training, Credentialing, and Hospital Privileging for Robotic Urological Surgery. *Journal of Robotics in Genitourinary Surgery*, Part 2:169-181.

Comparison of the Usability of Robotic Surgery Simulators

Alyssa Tanaka, M.S., Courtney Graddy, M.S., Roger Smith, Ph.D.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org, courtney.graddy@flhosp.org, roger.smith@flhosp.org

Haider M. Abdul-Muhsin, M.D.
Mayo Clinic
Scottsdale, AZ
Hma980@gmail.com

ABSTRACT

The introduction of simulation into minimally invasive robotic surgery is relatively recent and has seen rapid advancement; therefore, a need exists to develop training curriculums and to identify systems that will be most effective at improving surgical skills. Several robotic simulators have been introduced to support these aims, but their effectiveness has yet to be fully evaluated.

Currently, there are three simulators -- the daVinci Skills Simulator, Mimic dV-Trainer, and Surgical Simulated Systems' RoSS. While multiple studies have been conducted to demonstrate the validity of each system, no studies have been conducted which compare the value of these devices as tools for education and skills improvement.

This paper presents the results of an experiment comparing value, usability, and validity of all three systems. Subjects who were qualified as medical students or physicians (n=105) performed one exercise on each of the three simulators and completed two questionnaires, one regarding their experience with each device and a second regarding the comparative effects of the simulators. This data confirmed the face, content, and construct validity for the dV-Trainer and Skills Simulator. Similar validities could not be confirmed for the RoSS. Greater than 80% of the time, participants chose the Skills Simulator in terms of physical comfort, ergonomics, and overall choice. However, only 55% thought the skills simulator was worth the cost of the equipment. The dV-Trainer had the highest cost preference scores with 71% percent of respondents feeling it was worth the investment.

This work is the second component of a three-part analysis. In the previous study, the simulators were objectively reviewed and compared in terms of their system capabilities. The third part will evaluate the transfer of training effect of each simulator. Collectively, this work will offer end users and potential buyers a comparison of the value and preferences of robotic simulators.

ABOUT THE AUTHORS

Alyssa D.S. Tanaka, M.S. is a Systems Engineer at Florida Hospital's Nicholson Center. Her research work focuses on robotic surgery simulation and effective surgeon training. Her current projects include rapid prototyping of surgical education devices, the validation of a robotic surgical curriculum and evaluation of robotic simulation systems. She is a Modeling and Simulation PhD student at the University of Central Florida and previously earned a M.S. in Modeling and Simulation, Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida.

Courtney Graddy, M.S. is a Human Studies Research Coordinator at the Celebration Health Research Institute where she manages projects aimed at improving patient health outcomes, employee health, process improvement and simulation research. Her current projects focus on integrating technology into standard of care and evaluating its effects on patient health and patient satisfaction, as well as evaluating teaching modalities used to train surgeons. Her career began at the North Florida South Georgia Veterans Health System where she aided in the development of employee education materials and program planning and evaluation with the Geriatric Research Education and Clinical Center. She holds a Bachelors of Science in Health Education from the University of Florida and a Masters of Health Administration from the University of South Florida.

Haidar Mohamed Abdul-Muhsin, M.D. is an endourology fellow at Mayo Clinic in Scottsdale, AZ. Prior to this role, he was a urology fellow at the Global Robotics Institute at Florida Hospital at Celebration. Dr. Abdul-Muhsin has been practicing medicine for 10 years. His research interests include prostate cancer research with emphasis on robotic prostatectomy outcomes and research related to robotic surgical simulation. Currently, he is researching the contribution of laparoscopic surgical experience in developing robotic proficiency.

Roger Smith, Ph.D. is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading technology implementation. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *Innovation for Innovators: Leadership in a Changing World*. He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

Comparison of the Usability of Robotic Surgery Simulators

Alyssa Tanaka, M.S., Courtney Graddy, M.S., Roger Smith, Ph.D.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org, courtney.graddy@flhosp.org, roger.smith@flhosp.org

Haider M. Abdul-Muhsin, M.D.
Mayo Clinic
Scottsdale, AZ
Hma980@gmail.com

INTRODUCTION

Robotic surgery has introduced a new dimension into the surgical field. With the introduction of robotic technology between patient and surgeon, a need to master new skills has emerged. Medicine has come to the conclusion that the Halstedian training model (See one, do one, teach one) is no longer sufficient for teaching complex skills, especially robotic surgical skills (Cameron, 1997). A number of simulators have been developed to support training and skill assessment in robotic surgery. The currently available dedicated robotic simulators include: the da Vinci Skills Simulator (dVSS) by Intuitive Surgical Inc., also known as the "Backpack Simulator"; the dV-Trainer from Mimic Technologies Inc.; and the RoSS by Simulated Surgical Sciences LLC (Figure 1). The purpose of these simulators is to train surgeons prior to using the actual system and to allow them to acquire the necessary robotic skills to perform a safe surgery. All of these da Vinci simulators utilize a visual scene that is presented in a computer generated 3D environment providing challenging tests for practicing dexterity and machine operations. Originally, the simulated exercises trained basic robotic skills; however with advances in technology, surgeons can now train for specific procedures (e.g. nephrectomy and hysterectomy).

Figure 1. Simulators of the da Vinci robotic surgical system

Our hospital research laboratory has purchased each of these three simulators for the purpose of studying their effectiveness and applying them to the education of robotic surgeons, specifically for the Department of Defense (DoD). The DoD is interested in the effectiveness of the simulators to train military surgeons prior to and after returning home from deployments. This research is structured as three distinct stages.

From the first stage of this work, the authors summarized the objective characteristics of the three systems. This included descriptions of the exercises offered in each, metrics used to evaluate students, overview of the system administration functions, physical dimensions and configurations of the equipment, and comparisons of the costs of the devices and their support equipment (Smith & Truong, 2013). In the first simulator, the trainee sits at and operates the simulated environment using the actual da Vinci surgical console. The simulator is a custom computer appended to the surgical console through the actual surgical data port. While the simulator costs approximately

$100,000, the surgical console costs $500,000 incurring an investment of $600,000. Using this simulator, users can train using the actual hardware they would use during surgery; however, this requires the use of the surgical console that may be needed to conduct surgeries. Most hospitals may not have a dedicated training console, meaning that users would not have appropriate access to the simulator. The second is a standalone system that utilizes a graphic/gaming computer, connected to a custom desktop viewing and control device that replicates the hardware of the da Vinci surgeon's console. This system shares similar software with the dVSS, but does not require the use of any actual da Vinci hardware. The cost of this simulator is approximately $100,000. The third is composed of a completely customized replica of the da Vinci surgeon's console. Internally the simulator contains a graphic computer, a 3D monitor, and commercial Omni Phantom haptic controllers. This simulator uses unique software and is a little more than $100,000 (Smith &Truong, 2013).

This paper reports on the second stage of this research, in which the validity and usability of the simulators is examined. The third stage will be a measure of learning effectiveness using the systems.

Validity in Surgical Simulation

The validity of medical and surgical simulators is usually measured by the categories defined by McDougal (2007). This paper defines the most commonly recognized forms of validation as: *face, content, construct, concurrent, and predictive validity.* *Face validity* is typically assessed informally by users and is used to determine whether the simulator is an accurate representation of the actual system (i.e. the realism of the simulator). *Content validity* is the measure of the appropriateness of the system as a teaching modality. Experts who are knowledgeable about the device typically assess this via a formal evaluation. *Construct validity* is the ability of a simulator to differentiate between the performances of experienced users and those who are novices. *Concurrent validity* is the extent to which the simulator correlates with the "gold standard" and *predictive validity* is the extent to which the simulator can predict a user's future performance. Collectively, concurrent and predictive validity are known as criterion validity and are used as measures of the simulator's ability to correlate trainee performance with their real life performance. Face and content validity are most effective in evaluating the ability of a simulator to train a surgeon; however construct, concurrent, and predictive validity are most useful for evaluating the effectiveness of a simulator to assess a trainee.

The validity of all three simulators has been tested and reported separately for the da Vinci skill simulator (Hung, Zehnder, Patil, 2011; Kelly, Margules, Kundavaram, 2012; Liss, Abdelshehid, Quach, 2012), the dV-Trainer (Kenney, Wszolek, Gould, Libertino, Moinzadeh, 2009; Sethi, Peine, Mohammadi, 2009; Lee, Mucksavage, Kerbl, 2012) and the RoSS (Seixas-Mikelus, Kesavadas, Srimathveeravalli, 2010; Stegemann et al., 2013; Colaco, Balica, Su, 2012; Raza et al., 2013). To our knowledge only one publication has compared features of two of the simulators, but no comparative studies have been performed with all three of the systems (Liss MA, Abdelshehid C, Quach S., 2012). Thus, the current study aimed to compare all three commercially available da Vinci simulators and detail the findings for face, content, and construct validity for the three systems.

METHODS

Recruitment

Participants in this study included medical students, residents, fellows, and attending physicians. Participants were recruited from the University of Central Florida Medical School, courses held at the Nicholson Center, and two medical robotic conferences (World Robotics Gynecology Congress and Society of Robotic Surgeons Scientific Meeting). Subjects were excluded from participating if they indicated that they had participated in a formal robotic simulation-training course.

Each participant was categorized into one of three groups (i.e. Expert, Intermediate, or Novice) according to the self-reported number of robotic cases (i.e. procedures) he or she had performed. Individuals performing 0-19 robotic cases in which they had 50% or greater console time were categorized as Novices, individuals with 20-99 robotic cases were considered to be Intermediates, and individuals with 100 or more cases were considered to be Experts.

Materials

After being categorized into an experience level, each participant was assigned a specific order in which they used each of the simulators (Figure 2). This order system was used to identify and potentially eliminate any bias that may exist by using a specific system first. All participants completed one exercise on each of the simulators. The tasks chosen were Peg Board 1 in both the dV-Trainer and the dVSS and Ball Placement 1 in the RoSS. The same task was used for both the dV-Trainer and the dVSS because these systems share similar software and exercises. The RoSS software contains unique exercises and Ball Placement 1 is designed to teach the same skills as Peg Board 1.

Participant	Simulator Order
01	CBA
02	BAC
03	ACB
04	CBA
05	BAC
06	ACB
07	CBA
08	BAC
09	ACB

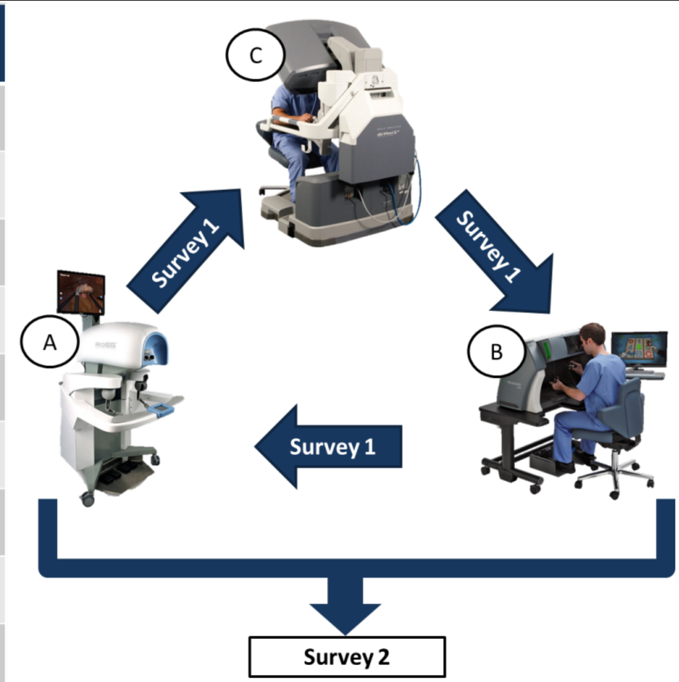


Figure 2. Rotating order of use by subjects, with survey order.

After each exercise on each simulator, participants completed a post questionnaire (Survey 1), which asked for feedback regarding their experience on that specific simulator. After using all three systems, subjects completed a second post questionnaire (Survey 2), which asked them to compare all three systems to each other. The participant's performance metrics were also collected from each of the simulators.

RESULTS

Demographics

Subjects were categorized as Novice (n=37), Intermediate (n=31), or Expert (n=37). Sixty-two percent of subjects were men and 38% were women with an average age of 43. On average, participants had 15 years in practice and 3 years of robotic experience. Seventy-six percent were attending physicians and 73% of participants were currently or had received robotic training, while 41% provided that they train residents and fellows. There were differences in the average age and number of years in practice of participants based on the classification of expert, intermediate or novice (number of robotic procedures). These are to be expected, since higher ages are required to achieve higher number of years of practice and larger numbers of robotic procedures.

Validation

The types of validity evaluated in this experiment were face, content, and construct. To analyze the systems for face validity and content validity, questions from Survey 1 were used. The questions were evaluated on a five point Likert scale (Strongly Disagree, Disagree, Neither Agree or Disagree, Agree, and Strongly Agree). Face validity was

analyzed by expert and intermediate feedback as recommended by Van Nortwick et al. (2010) because these are the users most familiar with the robotic system; however, only expert feedback was used for content validity because they have the best ability to judge the appropriateness of the system as a training tool. For construct validity, performance metrics such as Overall Score, Time to Complete, Number of Errors, and Economy of Motion were analyzed (Table 1).

Table 1. Questions and data used for different levels of validity.

Type of Validity	Evaluation	Type of Participant	Question/Metric
Face Validity	Survey 1	Expert and Intermediate	Q1: The hand controllers on this simulator are effective for working in the simulated environment (Likert).
			Q4: The device is a sufficiently accurate representation of the real robotic system (Likert).
Content Validity	Survey 1	Expert	Q2: The 3D graphical exercises in the simulator are effective for teaching robotic skills (Likert).
			Q5: The scoring system effectively communicates my performance on the exercise (Likert).
			Q6: The scoring system effectively guides me to improve performance on the simulator (Likert).
Construct Validity	Simulator	Experts and Novices	Overall Score (points)
			Number of Errors (count)
			Time to Complete (seconds)
			Economy of Motion (centimeters)

Face Validity

The responses of Intermediate and Expert participants (n=68) were used to determine face validity (Table 2). A Chi-square test of independence was used to evaluate the distribution of scores for a specific simulator in relation to the order of the system's presentation to the subject. This analysis indicated that there was no difference in participants' answers according to the order in which the systems were presented; and established that no bias was present due to the presentation order ($p>0.05$). These questions asked participants to evaluate whether the hand controllers on the simulator were effective for working in the simulated environment (Question 1) and if the device is a sufficiently accurate representation of the real robotic system (Question 4). For both questions, the RoSS had the lowest average score, dV-Trainer had the second highest score, and the dVSS had the highest score of the three. A repeated measures ANOVA verified that the systems were scored differently for both questions ($p<0.001$).

Table 2. Average scores from a 5-point Likert scale on face validity.

	DVSS	dV-Trainer	RoSS
Q1: The hand controllers on this simulator are effective for working in the simulated environment.	4.80	3.62	2.17
Q4: The device is a sufficiently accurate representation of the real robotic system.	4.65	3.45	1.82

Content Validity

Expert (n=34) responses were used to determine whether the simulators were appropriate teaching modalities (Table 3). As seen in Table 3, 100% of participants either agreed or strongly agreed that the 3D graphical exercises in the dVSS were effective for teaching robotic skills while 59% disagreed or strongly disagreed that the RoSS' capabilities were effective. When asked if the scoring system effectively communicated their performance, 88% of dVSS users agreed or strongly agreed, while 79% of dV-Trainer users agreed or strongly agreed. Similarly, 91% and

82% of participants agreed or strongly agreed that the dVSS and dV-Trainer, respectively, effectively guided them to improve their performance, while only 36% felt the RoSS provided the same guidance.

Table 3. Scores on a 5 point Likert scale for content validity questions.

Likert Score	Strong Dis	Disagree	Neither	Agree	Strong Agree
Q2: *The 3D graphical exercises in the simulator are effective for teaching robotic skills.*					
DVSS	0%	0%	0%	35.3%	64.7%
dV-Trainer	2.9%	5.9%	11.8%	50.0%	29.4%
RoSS	20.6%	38.2%	17.6%	17.6%	5.9%
Q5: *The scoring system effectively communicates my performance on the exercise.*					
DVSS	2.9%	5.9%	2.9%	38.2%	50.0%
dV-Trainer	2.9%	2.9%	14.7%	55.9%	23.5%
RoSS	17.6%	20.6%	26.5%	29.4%	5.9%
Q6: *The scoring system effectively guides me to improve performance on the simulator.*					
DVSS	0%	0%	8.8%	61.8%	29.4%
dV-Trainer	2.9%	2.9%	11.8%	61.8%	20.6%
RoSS	18.2%	18.2%	27.3%	33.3%	3.0%

Construct Validity

The overall score, number of errors, time to complete, and economy of motion scores collected by the simulators for Experts (n=37) and Novices (n=37) were used to compare construct validity (Table 4). Overall score is a metric synthesized by multiple metrics and is specific to the individual simulator. Intermediate subjects were not included in the construct validity analysis because it was only necessary to look if the simulator could distinguish specifically between novice and expert users.

For the RoSS, the analysis has 23 missing data points because the system does not report scores when a user exceeds a maximum exercise time or chooses to terminate the exercise before completion. This resulted in a sample of 30 experts and 21 novices on that system. A Mann-Whitney U test showed that the distributions of time ($p=0.221$), number of errors ($p=0.644$), and economy of motion ($p=0.566$) were not statistically different for the experts compared to the novice group. The overall score metric is not automatically exported by the simulator and therefore was not analyzed for this sytem.

The dV-Trainer analysis of experts (n=37) and novices (n=37) had three missing values for economy of motion and completion time and five for the overall score metric, thus the analysis contained varying number of subjects. A Mann-Whitney U test showed that the distribution of the overall scores was not significantly different for the expert compared to the novice group ($p=0.061$). These tests did confirm statistical differences for economy of motion ($p<0.001$) and time to complete ($p<0.001$) for this system with a lower economy of motion value and shorter completion time for expert users compared to novices.

The dVSS analysis included all novice (n=37) and expert (n=37) participants. Using a Mann-Whitney U test, time to complete ($p<0.001$) and overall score ($p=0.006$) were significantly different for the expert compared to the novice group. The expert group had a higher score and a shorter completion time compared to the novice group. However, economy of motion did not show a statistical difference with this analysis ($p=0.216$).

Table 4. Mann-Whitney U test level of significance on construct validity measures

	DVSS	dV-Trainer	RoSS
Time to Complete	p<0.001	p<0.001	p=0.221
Overall Score	p<0.01	p=0.061	n/a
Economy of Motion	p=0.216	p<0.001	p=0.566
Number of Errors	n/a	n/a	p=0.644

The construct validity of the simulators was more specifically analyzed in terms of the self-reported number of cases of all participants (n=105) using a non-parametric correlation coefficient (Spearman's). For the RoSS, 30 participants were excluded from the analysis. For the participants that were included in the analysis (n=75), there was not a significant correlation between time to complete (p=0.181), number of errors (p=0.563), or economy of motion (p=0.390) with the total number of robotic cases performed.

For the dV-Trainer, four participants were excluded from the entire analysis and two participants were excluded from the overall score (Overall Score n=99; Economy of Motion and Time to Complete n=101). When analyzing the number of participants' robotic cases, there was a statistically significant correlation between overall score (p=0.03), economy of motion (p<0.01), and time to complete (p<0.01). The correlation value was negative for economy of motion and time to complete, showing that with a greater number of robotic cases, the time taken and distance moved decreased. The correlation was positive for overall score indicating that the participants' score increased with the number of robotic cases performed.

For the dVSS, two participants were excluded from the analysis (n=103). When analyzing the metrics in terms of the total number of robotic cases performed, there was a statistically significant difference between overall score (p =0.01) and time to complete (p <0.01). The correlation value was negative for time and positive for overall score, signifying that with more robotic cases the time taken decreased and the score increased. There was not a statistically significant correlation between economy of motion and the total number of robotic cases performed (p=0.105).

Table 5. Correlation between level of experience and simulator scores

	DVSS	dV-Trainer	RoSS
Overall Score	p=0.001	p=0.031	n/a
Time to Complete	p<0.001	p<0.001	p=0.181
Economy of Motion	p=0.105	p<0.001	p=0.390
Number of Errors	n/a	n/a	p=0.563

Usability (Preference)

The questions from the Survey 2 were used to understand the preference of the subjects when using the simulators. All subjects were included in this analysis except for two participants who were dropped from the analysis because they did not complete the questionnaire. The participant's responses to the usability questions can be seen in Figure 3:

- *If you are (were) a program director, which simulator would you choose for your trainees;*
- *In which simulator were you physically more comfortable;*
- *Which simulator had the best hand controls;*
- *Which simulator had the best foot controls;*
- *Which simulator had the best 3D vision;*
- *Were you feeling stressed or annoyed by any of the simulators?*

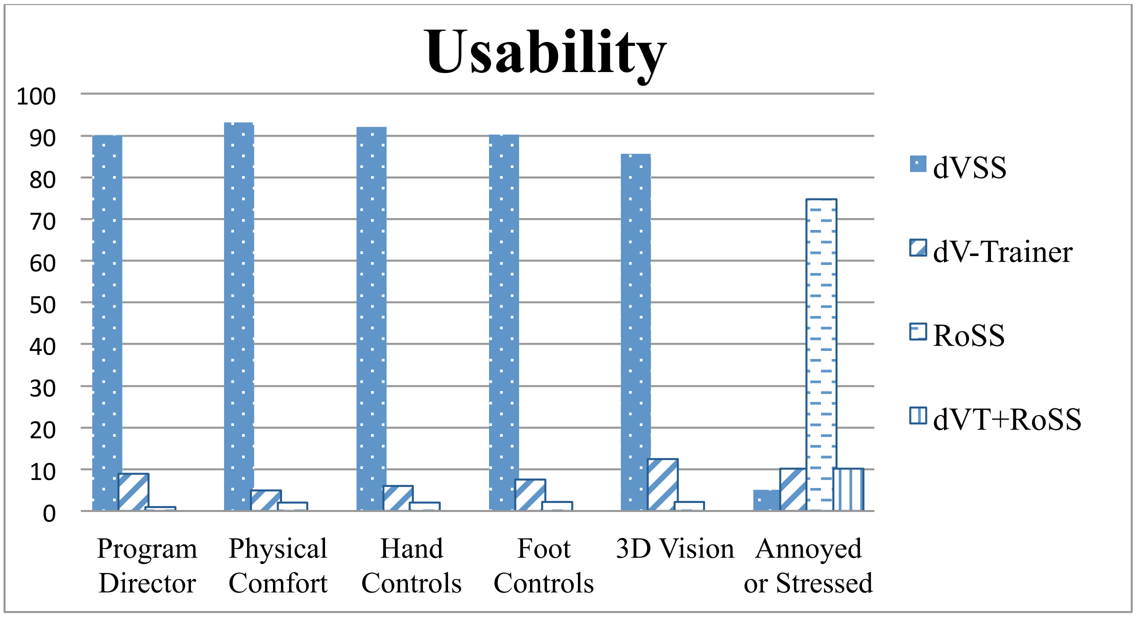


Figure 3. Description of usability responses

Overall, most participants preferred the dVSS and indicated that they would choose this device as a training system if they were a program director. Participants not only felt most comfortable in the dVSS, but also felt that the system had the best control and vision equipment. The least preferred system was the RoSS which most participants also agreed made them feel stressed or annoyed. Ten percent of participants also responded that they felt stressed or annoyed by both the dV-Trainer (dVT) and the RoSS.

Cost

All participants were also asked to provide feedback on their simulator preference in terms of the cost of the system. The responses were analyzed in terms of the frequency of the responses given. Most participants felt that the mimic dV-Trainer was worth the investment; while most felt that the RoSS was not worth the money. When asked about the dVSS, only 56% of participants agreed that it was worth the investment. Figure 4 provides a full description of the responses.

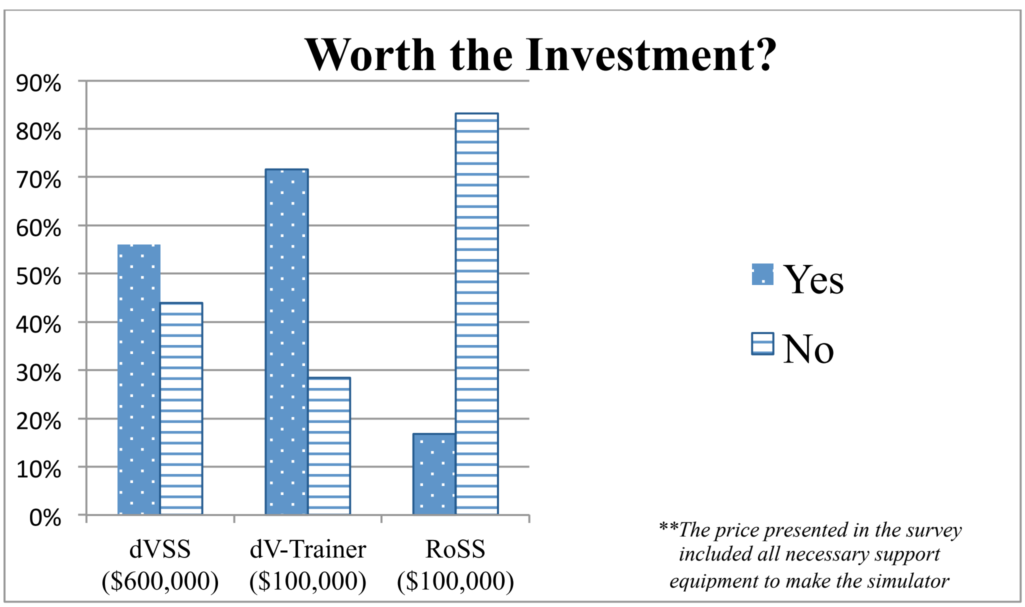


Figure 4. Description of cost preferences

DISCUSSION

The aim of this study was to conduct a comparison of the three commercially available simulators used to train surgeons on the daVinci robotic system. The study was performed for the US Army to assist them in making a purchasing and deployment decision regarding robotic simulators. Their interest is in re-training robotic surgeons who have been deployed to combat zones, where they have served as trauma surgeons for many months. Prior to resuming their robotic specialties, these surgeons need a program to both refresh and re-validate their robotic skills. This study provided information about the face, content, and construct validity as well as usability of the systems. The simulators were perceived to be different in their representation of the real robotic system. The dVSS was most preferred in terms of ergonomics and usability; however, most participants did not feel that this system was worth a $600,000 investment. In terms of cost, most participants agreed that the dV-Trainer had the best cost-effectiveness. The RoSS was the least preferred system for comfort and other usability aspects (i.e., hand controls, foot controls, and 3D interface), with most participants feeling stressed or annoyed when using the system. This study was unable to validate the face, content, or construct validity for this system.

The dVSS leverages the actual hardware used to perform robotic surgeries for use in the simulated environment, which allows for a more realistic experience, but decrease its availability and creates a higher cost for training than other robotic simulators. Economy of motion was not able to differentiate novices from experts in the dVSS, which could be attributed to the ease of use of the controllers allowing novices to move the controls as efficiently as experts. The generous workspace of the dVSS could also have an impact on the lack of difference. In contrast to the dVSS, the dV-Trainer is a standalone simulator and does not require the support of the daVinci hardware to operate. This allows for better accessibility and requires less of an investment for training. The overall score aspect of construct validity may not have shown a difference between novices and experts because of the way that the scoring is developed. The scoring system is constructed with a “ceiling” that prevents users from achieving a high overall score without attaining high scores across multiple metrics.

Currently, there is limited data available that confirms construct validity of the RoSS. Similarly to Raza (2013), this study was unable to confirm a difference between experts and novices in terms of time taken to complete the exercise. Time to complete, as well as economy of motion, is considered a highly relevant measurement of expertise levels for robotic surgeons (Perrenot, Perez, Tran, Jehl, Felblinger, Bresler, & Hubert, 2012). To our knowledge this three-part study is the first to compare all three available systems. This study involved the largest sample size and diversity of participants (i.e., experience levels, number of robotic cases, and subspecialty type) thus far in relevant publications. The lack of consistency in the available exercises and scoring systems across the three systems was a limitation to the study. Considerations for future research would be to use more complex exercises and increase the depth of the face and content validity evaluation.

Current research is focused on the effectiveness of the simulators and objectively measuring the transfer of training to the actual robotic system. All three simulators will be examined in this final stage of the experiment; however, the results of this three-part study will guide the choice of simulators used for future studies at Florida Hospital Nicholson Center and may also influence decisions at other laboratories. Also, this research may impact the purchasing decisions of customers for these devices.

REFERENCES

Cameron, J. L. (1997). William Stewart Halsted. Our surgical heritage. *Annals of surgery*, *225*(5), 445.

Colaco M, Balica A, Su D. (2012) Initial experiences with RoSS surgical simulator in residency training: a validity and model analysis, J Rob Surg, 7(1):71-75

Hung AJ, Patil MB, Zehnder P. (2012). Concurrent and predictive validation of a novel robotic surgery simulator: a prospective, randomized study. J Urology, 187(2):630-7.

Hung AJ, Zehnder P, Patil MB. (2011). Face, content and construct validity of a novel robotic surgery simulator. J Urology, 186(3):1019-24.

Kelly DC, Margules AC, Kundavaram CR. (2012). Face, content, and construct validation of the da Vinci Skills Simulator. J Urology, 79(5):1068-72.

Kenney P, Wszolek M, Gould J, Libertino J, Moinzadeh A. (2009). Face, Content, and Construct Validity of dV-Trainer, a Novel Virtual Reality Simulator for Robotic Surgery. J Urology, 73(6):1288-92.

Lee JY, Mucksavage P, Kerbl DC. (2012) Validation study of a virtual reality robotic simulator--role as an assessment tool? J Urology, 187(3):998-1002.

Liss MA, Abdelshehid C, Quach S. (2012). Validation, Correlation, and Comparison of the da Vinci Trainer(™) and the da Vinci Surgical Skills Simulator(™) Using the Mimic(™) Software for Urologic Robotic Surgical Education. J Endourol, 26(12):1629-34.

McDougall EM. (2007). Validation of surgical simulators. J Endourol, 21(3):244-7.

Sethi AS, Peine WJ, Mohammadi Y. (2009). Validation of a novel virtual reality robotic simulator. J Endourol, 23(3):503-8.

Perrenot, C., Perez, M., Tran, N., Jehl, J. P., Felblinger, J., Bresler, L., & Hubert, J. (2012). The virtual reality simulator dV-Trainer® is a valid assessment tool for robotic surgical skills. *Surgical endoscopy*, *26*(9), 2587-2593.

Raza SJ, Chowriappa A, Stegemann A, Ahmed K, Shi Y, Wilding G, Kaouk J, Peabody J, Menon M, Kesavadas T, Guru KA. (2013). Construct Validation of the Fundamental Skills of Robotic Surgery (FSRS) Curriculum. Podium presentation, AUA Annual Congress, San Diego,

Seixas-Mikelus SA, Kesavadas T, Srimathveeravalli G. (2010). Face validation of a novel robotic surgical simulator. J Urology, 76(2):357-60.

Smith, R., & Truong, M. (2013, January). Robotic Surgical Education with Virtual Simulators. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (Vol. 2013, No. 1). National Training Systems Association.

Stegemann A, Kesavadas T, Rehman S, Sharif M, Rao M, duPont N, Shi Y, Wilding G, Hassett J, and Guru K. (2013). Development, Implementation, and Validation of a Simulation-Based Curriculum for Robot-Assisted Surgery. J Urology, 81(4):767-74.

Van Nortwick, S. S., Lendvay, T. S., Jensen, A. R., Wright, A. S., Horvath, K. D., & Kim, S. (2010). Methodologies for establishing validity in surgical simulation studies. *Surgery*, *147*(5), 622-630.

From Design to Conception: An Assessment Device for Robotic Surgeons

Alyssa Tanaka, M.S.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org

Manuela Perez, M.D.
University Hospital of Nancy
Nancy, FR
m.perez@chu-nancy.fr

Mireille Truong M.D., Khara Simpson M.D.
Columbia University Medical School
New York, NY
Mireille.truong@gmail.com, kmsimpmd@yahoo.com

Gareth Hearn, Roger Smith, Ph.D.
Florida Hospital ISA, Florida Hospital NC
Orlando, FL, Celebration, FL
Gareth.hearn@flhosp.org, roger.smith@flhosp.org

ABSTRACT

The daVinci Surgical System offers surgeons improved capabilities for performing complex minimally invasive procedures; however, there is no standardized assessment of robotic surgeons and a need exists to ensure that a minimal standard of care is provided to all patients. The Department of Defense and governing surgical societies convened consensus conferences to develop a national initiative, resulting in a curriculum called the Fundamentals of Robotic Surgery (FRS). FRS is comprised of an online curriculum and a psychomotor skills dome.

This paper describes the production process used to create a psychomotor skills assessment device - the FRS Dome. The device was designed to measure the essential skills that are required of any robotic surgeon and to provide a basis upon which to grant or deny privileging with the robot. It was constructed to test seven tasks of manual dexterity: Docking, Ring Tower Transfer, Knot Tying, Suturing, 4th Arm Cutting, Puzzle Piece Dissection, and Energy Dissection.

The initial design of the device was created by a committee of experienced minimally invasive surgeons, with a background in testing protocols and materials. The design was rendered in computer animation, which kick-started a prototyping effort with physical materials. These included platinum cure silicone approximating human tissue and a 3D polyjet printer for the structural framework. Usability testing was conducted and iterative modifications were made to improve ergonomics, standardization, and cost requirements. Final CAD diagrams and specifications were created and distributed to medical and simulation companies for both physical and digital manufacturing. This development process demonstrates the evolution of a simulation and a physical testing device based on international expert consensus. The specifications are open source, allowing competitive production and future iterations. The goal of this paper is to discuss how this device evolved from an idea to a manufactured product and a digital simulation.

ABOUT THE AUTHORS

Alyssa D.S. Tanaka, M.S. is a Systems Engineer at Florida Hospital's Nicholson Center. Her research work focuses on robotic surgery simulation and effective surgeon training. Her current projects include rapid prototyping of surgical education devices, the validation of a robotic surgical curriculum and evaluation of robotic simulation systems. She is a Modeling and Simulation PhD student at the University of Central Florida and previously earned a M.S. in Modeling and Simulation, Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida.

Manuela Perez, M.D. is a practicing General Surgeon at the University Hospital of Nancy-France, where she also serves as an Assistant Professor in General Surgery and Anatomy. Dr. Perez has been practicing medicine for 14 years and graduated with her PhD in Robotic Surgery, with a thesis entitled "Telesurgery: From Training to Implementation." Currently, she is working as a Research Fellow at the Florida Hospital Nicholson Center and working under a grant from the Department of Defense researching various aspects of Telesurgery.

Mireille Truong, M.D. is a Minimally Invasive Gynecology Fellow at Columbia University, where she serves as an assistant attending and clinical instructor of Obstetrics and Gynecology (OB/GYN). During her residency in OB/GYN at the University of Illinois at Chicago, she received a number of awards and honors, including Administrative Chief Resident, the American Association of Gynecologic Laparoscopists' Special Resident in Minimally Invasive Gynecology Award and Best Overall Excellence in Gynecologic Care Award. She has dedicated her time to education of medical students and physicians via an array of academic and teaching appointments, professional organization board positions, peer-reviewed publications and presentations at various national and international conferences. Her research interests include surgical education, robotic simulation and minimally invasive gynecologic surgery.

Khara Simpson, M.D. is a second year fellow in minimally invasive surgery at Columbia University, where she serves as an assistant attending and instructor of obstetrics and gynecology. She completed her medical school education at Howard University College of Medicine where she was inducted into the Alpha Omega Alpha honor medical society. Following, she completed her OB/GYN residency at Johns Hopkins University and served as administrative chief resident. She recently completed a one year research fellowship at the Florida Hospital Nicholson Center focusing on robotic surgery simulation. Her additional research interests include resident education and simulation training, and best practices to promote cost effective care.

Gareth Hearn a Mechanical Engineer at the Institute for Surgical Advancement, part of Florida Hospital Orlando. He is responsible for the Prototype Design Lab which is focused on minimally invasive surgical devices and positioned to accelerate the product development cycle. He has 8 years' experience working in the Dept. of Defense training and simulation industry. He is pursuing a Master's in Systems Engineering at the University of Florida and has previously earned a B.S. in Mechanical Engineering from UF.

Roger Smith, Ph.D. is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading technology implementation. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *Innovation for Innovators: Leadership in a Changing World*. He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

From Design to Conception: An Assessment Device for Robotic Surgeons

Alyssa Tanaka, M.S.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org

Manuela Perez, M.D.
University Hospital of Nancy
Nancy, FR
m.perez@chu-nancy.fr

Mireille Truong M.D., Khara Simpson M.D.
Columbia University Medical School
New York, NY
Mireille.truong@gmail.com, kmsimpmd@yahoo.com

Gareth Hearn, Roger Smith, Ph.D.
Florida Hospital ISA, Florida Hospital NC
Orlando, FL, Celebration, FL
Gareth.hearn@flhosp.org, roger.smith@flhosp.org

INTRODUCTION AND BACKGROUND

Robotic surgery has been established as an innovative approach in surgery due to a telemanipulator device, which introduced a new dimension into surgical tools. This device allows surgeons to manipulate robotic arms from a remote console to perform complex surgical procedures. Robotic surgical systems overcome laparoscopic limitations and facilitate the performance of minimally invasive surgery due to 3D vision, 7-degree-of-freedom instruments, tremor abolition, motion amplification, and stabilization of the camera (Patel et al., 2013; Hubens, Coveliers, Balliu, Ruppert, & Vaneerdeweg, 2003; Blavier, Gaudissart, Cadière, & Nyssen, 2007). The system also offers 10x magnification, wristed instruments, and a third working arm. Currently, the only system is Intuitive's da Vinci Surgical System (Figure 1).

Figure 1. da Vinci Surgical System

Robotic surgery has demonstrated safety and effectiveness for urologic, gynecologic, ENT, and complex general surgery procedures (Barbash, Friedman, Glied, & Steiner, 2014; Serati et al., 2014; Maan, Gibbins, Al-Jabri, & D'Souza, 2012; Luca et al., 2013; Zureikat et al., 2013). Exponential growth of minimally invasive procedures, particularly robotic-assisted procedures, raises the question of how to assess robotic surgical skills. This device also introduces a specific need for training and certification to ensure a minimal standard of care for all patients. Some institutions have attempted to develop and validate robotic training in regards to specific specialties (Chitwood et al., 2001; Geller, Schuler, & Boggess, 2011; Grover, Tan, Srivastava, Leung, & Tewari, 2010; Chowriappa et al., 2014; Jarc & Curet, 2014); however, the lack of a national standard has pushed surgical societies (e.g. the Society of American Gastrointestinal and Endoscopic Surgeons and Society of Robotic Surgery) to develop a unified approach and standard for robotic skills training (Zorn et al., 2009).

To develop a comprehensive model for robotic surgery, the Department of Defense, Veterans Administration, and fourteen surgical specialty societies convened multiple consensus conferences to create the Fundamentals of Robotic Surgery (FRS) curriculum. A similar education and training initiative was implemented for use in laparoscopic surgery, which resulted in the Fundamentals of Laparoscopic Surgery (FLS). FRS Conference participants included more than 80 subject matter experts (SMEs), consisting of surgeons, psychologists, engineers, simulation experts, and medical educators (Smith, Patel, Chauhan, & Satava, 2013).

The committee's vision of FRS was driven by two main goals: to ensure a perfect understanding of the basics of robotic surgery and to develop a psychomotor skills program that focused on basic robotic tasks. The intended users for this program are novice robotic surgeons, who could be residents or fellows and attending surgeons

who have never used the robotic system. The committee began by outlining outcomes measures and metrics, which touched on the essential cognitive, psychomotor, and team training skills. This resulted in a prioritized matrix of 25 robotic surgery concepts, which is the core material used in the design and development of the FRS Curriculum (Smith, Patel, Satava R, 2013). Two assessment tools were created: an online curriculum for knowledge and team training skills and a device for psychomotor skill training and evaluation (Levy, n.d.).

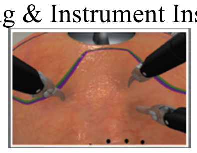
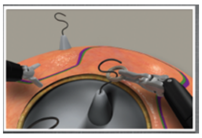
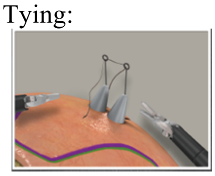
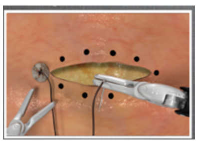
This paper discusses the process for designing and creating the physical device, known as the FRS dome. The purpose is to share the evolution of an idea to a usable device. The dome was conceived by experts who identified a clear need for robotic education and collectively developed a solution to fill the gap. The medical field is a constant progression of new concepts, devices, and technology. This paper also outlines the framework for which others can develop and introduce new concepts in medicine and other domains.

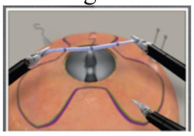
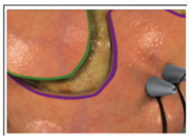
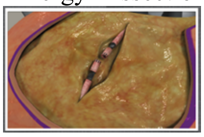
BRAINSTORMING AND CONCEPT DEVELOPMENT

Exercise Development

Of the 25 FRS concepts, 16 are directly linked with psychomotor skills. The FRS committee members then identified seven exercises that incorporated all 16 skills. These exercices include docking and instrument insertion, tower transfer, knot tying, railroad track, 4th arm cutting, puzzle piece dissection, and vessel energy dissection (Table 1). *Docking and instrument insertion* is an essential and unique robotic skill to begin a procedure. Failure at this stage of the procedure can compromise the surgery. Ring *Tower transfer* is a non-surgical exercise that introduces the utilization of endowrist manipulation and the 7 degrees of freedom to surgeons. *Knot tying* and *railroad track* are the base of a suturing exercise. The technology introduced in the wristed instruments facilitates the performance of these tasks. *4th arm cutting* is another task specific to robotics, which tests surgeon's autonomy. The 4th arm allows surgeons to manage three instruments by using a foot pedal to switch between working arms. *Puzzle piece* and *vessel energy dissection* are critical tasks, which incorporate complex articulation of instruments and application of energy (i.e. cauterization and cutting).

Table 1: Description of the basic psychomotor skills attached to the seven FRS tasks.

Exercises	Skills
Task1: Docking & Instrument Insertion: 	- Docking - Instrument insertion - Eye-hand coordination - Operative field of view
Task 2: Ring Tower Transfer: 	- Eye-hand coordination - Camera navigation - Clutching - Wrist articulation - A-traumatic handling
Task 3: Knot Tying: 	- Knot tying - Suture handling - Eye-hand coordination - Wrist articulation
Task 4: Railroad Track: 	- Needle handling & manipulation - Wrist articulation - A-traumatic handling - Eye-hand coordination

Task 5: 4th Arm Cutting: 	- Multiple arm control & switch - Cutting - A-traumatic handling - Eye-hand coordination
Task 6: Puzzle Piece Dissection: 	- Sharp and blunt dissection - Cutting - A-traumatic handling - Eye-hand coordination - Wrist articulation
Task 7: Vessel Energy Dissection: 	- Energy sources use - Sharp dissection - Cutting - Multiple arm control - A-traumatic handling - Eye-hand coordination

Device Development

The FRS committee envisioned all of the exercises contained on the outer surface of a single device. This would allow for the exercises to be administered quickly and easily, incur less cost, and ensure uncomplicated storage and transportation. The semi-spherical form (i.e. the dome), was quickly decided on as a shape which would integrate with the current robotic system. They depicted their ideas through simple drawings and crude models made from materials found on hand. During initial design planning, conference participants experimented with a variety of arrangements of the exercises on the dome.

A final sketch was developed and delivered to a 3D digital artist to create static pictures of the device, along with an animation of the performance of each exercise. The CGI provided the first formal images of the dome, which gave life to the device and proved feasibility. The realistic animations showed the exercises being performed and gave committee members a visual concept of how the device would function (Figure 2).

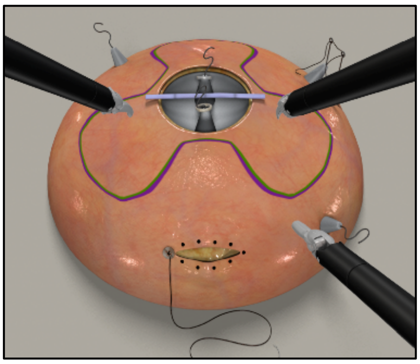


Figure 2. The initial 3D graphic FRS dome design

PROTOTYPING

The prototyping process began using the ideas developed in the design meeting and the CGI. This process would prove to be fundamental in confirming the design expectations. It was essential to determine if a single device could physically house all of the exercises effectively, if the planned architecture was compatible with the robotic system, and if the outcomes of the exercises could be measurable and reproducible.

Low-fidelity Prototypes

Low-fidelity prototypes (LFPs) were created using simple and inexpensive materials. None of the materials used in the LFPs were intended for inclusion in a final product. These materials were chosen because they were readily available, inexpensive, and easy to manipulate to test fit and function. These materials allowed rapid trial

and error testing of the technical aspects, clarifying requirements, and proving usability. The testing of the LFPs was performed using the da Vinci Surgical System and was video recorded. These recordings were sent to FRS committee members to provide their feedback. Each LFP resulted in multiple improvements to the designs, which were tested on subsequent prototype versions.

The base model of the LFPs was created using half of an 8" Styrofoam sphere as the support structure, yellow felt material as the fat layer, a latex swimming cap for the skin layer, and straws for the embedded vessels. The base of the towers was constructed using synthetic foam blocks carved into a cone shape (Figure 3). The exercise patterns were drawn onto the surface using a permanent marker.

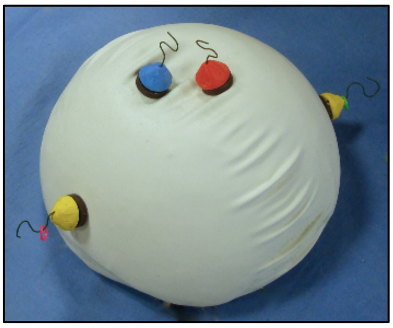


Figure 3. Base of Low Fidelity Prototypes

The LFPs evolved over six iterations, all of which introduced design improvements (Figure 4). At the earliest phase in LFP testing, it was quickly realized that the dome size was too large to fit under the robot arms appropriately. So, the dome size was decreased from 8" to 7". Another modification made early in the LFP development was to change the 4th arm cutting band from a rigid tube to an elastic band. This allowed for the user to adequately stretch the band prior to each cut.

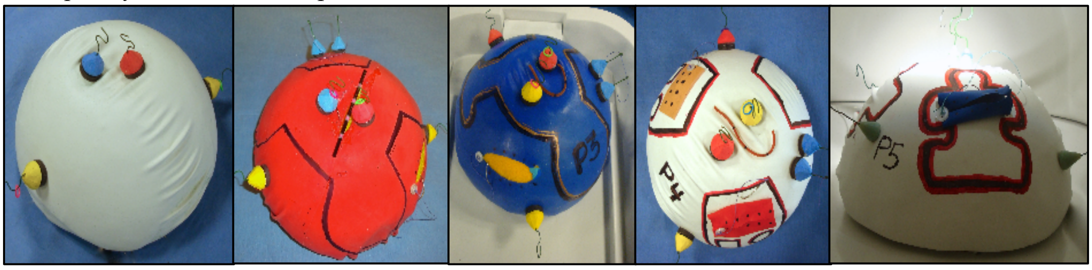


Figure 4. Iterations of LFPs

The suturing and dissection exercises involved the most modifications during the LFP stages. The original cloverleaf shape, used for the dissection exercise, was found to be too large and did not allow for the surgeons to access the section of the shape that was located on the backside of the dome. The size of the pattern was reduced; however, this did not mitigate the accessibility issue. The team experimented with other options, such as splitting the clover leaf into three sections and adding smaller shapes to the center of the cutting area. This design was not practical because once the smaller shapes were cut, the latex receded and inhibited surgeons from cutting the surrounding shape.

Eventually, the dissection shape evolved to a puzzle piece that incorporated all of the prerequisites for the dissection exercise (i.e. an accessible shape and a complex design). By using this compact pattern it became clear that all exercises could be grouped into an area covering only one third of the surface of the dome. This opened the opportunity to replicate the cluster of exercises three times on the surface, reducing the materials and costs for repeatedly practicing with the device. Another obstacle was to build the suturing exercise with the adequate materials and placements, to ensure a realistic feeling of suturing. Originally, the incision was made into the latex swim cap, however the latex would tear away and recede after the incision was cut in this model. Two versions of the suture module were experimented with: an embedded silicone and an external latex model. Eventually the embedded silicone model was chosen as the most realistic and practical for the exercise. Ultimately, the basic structural changes found in the low-fidelity prototyping were:

- The dome base needed to be reduced to 7"
- The dome base needed to be substantial in weight to keep from moving under the force of the robot
- A smaller, yet equally complex dissection shape was necessary
- The exercise sets could be grouped to allow them to be repeated on the surface of a single dome

- The magnets which held the towers to the dome needed to be of sufficient strength to hold through the layers of fat and skin

High-Fidelity Prototypes

The high-fidelity prototypes (HFPs) were made using higher quality, custom materials. These materials had the desired qualities of the final product and could be used as a basis for the large scale manufacturing process. The styrofoam base from the LFPs was replaced with a support structure that was printed using a 3D polyjet printer (Figure 5). A polyjet type 3D printer works similarly to an inkjet printer in that it distributes layers of polymer to build the desired design, which is cured by UV light. This type of printer was chosen because of the versatility allowed by printing multiple materials at once. Also, the jet lays $16\mu m$ layers of liquid polymer, which gives printed parts a finer resolution. Using this printer, a dome shell with a lid was created. The shell and lid had divots covering the surface, allowing for magnets to be moved to many different placements on the dome during design experiments. A small jig was also created using the 3D printer. Prior to the creation of the jig, the wires were made by hand, but the jig enables the standardized creation of the S-shaped and I-shaped tower wires. The price to print these items was approximately $1,000.

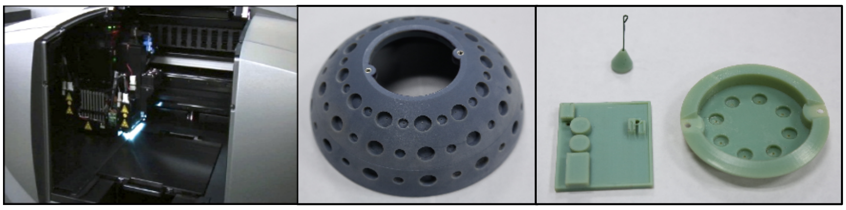


Figure 5. 3D printer with 3D printed dome, cap, towers, and jig

The synthetic tissue layers were created using Smooth-On platinum cure silicone products. These are two part silicones, which can be colored and mixed with other additives to achieve the desired product attributes such as durometer. The silicone used for the "fat" layer gave a gel-like and slightly sticky texture (Eco-flex Gel), while the "skin" silicone had a more firm and non-sticky quality (Ecoflex-0030). These silicones were chosen because they gave the closest resemblance to actual tissue properties. The fat silicone was poured directly onto the dome to the desired thickness. A clay mold was then made to replicate that thickness, which was used to form the skin layer (Figure 6). Embedded in the skin was a layer of polyester mesh, which helped to provide structure and stability of the skin. Small vessels were also created by quickly curing the silicone to a small tube. Using these materials we were able to create a set of synthetic tissues for less than $20.

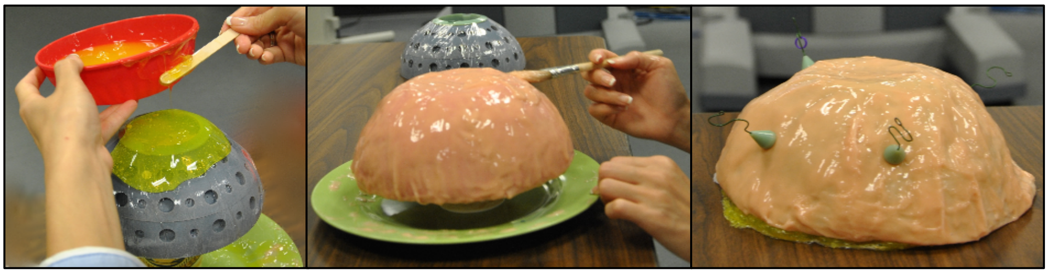


Figure 6. Pouring of silicones and first HFP

The puzzle piece shape and the other markers were drawn on the skin surface using a permanent marker. The exercises were drawn on in different locations, sizes, and orientations for the first HFP. After testing the HFP on the robotic system, we finalized the size and orientation of the exercises on this new dome. This is important because as learned in the LFP stage, the exercises needed to be placed strategically to compensate for the range of movement of the robotic arms. Despite having 7 degree-of-freedom instruments, there are still limitations to the amplitude of the movement of the robotic arms. We also determined that three trials of each exercise could fit on one dome, so each work station (i.e. group of exercises) repeated at 120 degree increments on the dome. Eventually, we determined that after dissecting the three vessels significant space was available for more dissection in the fat layer. So, we added three additional vessels located to the right of the original vessels and out of range of potential damage from other exercises (Figure 7). By doing so, the fat could be used six times and the skin used three times, which incurs lower costs for the materials used during training.

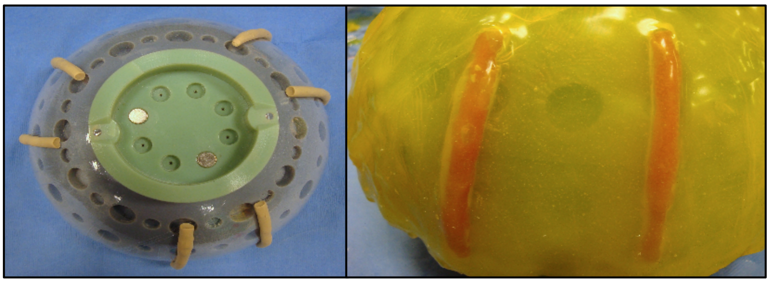


Figure 7. Vessel placement on dome and in fat

Over many iterative models, we improved our techniques and experimented with different materials and additives to achieve the desired qualities. For example we began adding a Thixotropic additive to thicken the mixture and allow us to cast the material onto a curved surface. We also tested different inks and techniques of printing the shapes and markers on the skin; however, most inks and paints cannot be used on silicone. We decided to use a silicone based paint product, which cured the design to the silicone surface.

We 3D printed miniature dome models (2” in diameter) to begin testing molding materials. We created silicone molds and used a urethane plastic to cast the model. By doing this we realized that the original 3D printed material was porous and caused bubbling in the molding, leading to surface bubbles on casted models. So, a new full sized dome was printed in a smoother and less porous material, which would be better for manufacturing. The new dome shell and cap was designed with divots only at the locations necessary for holding a tower (Figure 8).

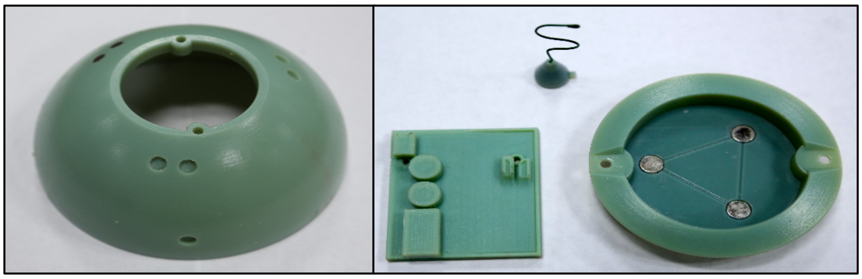
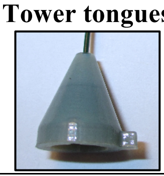
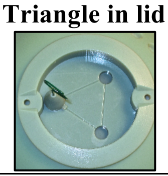
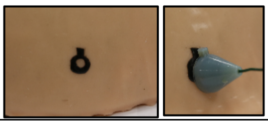
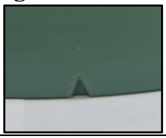
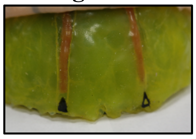
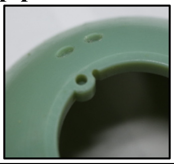


Figure 8. Final 3D printed dome shell

Since this device will be used for training and education, a high level of standardization is necessary. For this we added small markers that ensure the pieces are assembled correctly and in a standardized manner for all participants. Table 3 details the standardization pieces.

Table 3. Description of the Standardization Markers

Standardization Markers	
Tower tongues 	Used to orient the towers in the correct direction for each exercise.
Triangle in lid 	Used to show proper orientation of the towers that are placed in the cap. The towers are placed in the two locations directly in line with the puzzle piece and with the tower tongues on the corresponding line of the triangle. This ensures that the S-shaped towers face the correct direction for all users.
Tower orientation markers 	These markers are used to show the placement of the towers on the skin and the orientation of the tower. The towers are placed on the marker with the tongue aligned with the tongue mark. This ensures that all towers face the correct way.

Triangles on dome shell 	These small markers are located at 120 degree increments on the lower edge of the dome. They signify where the embedded vessels should be located when the tissue layers are placed on the shell.
Triangles on fat 	There are two types of triangle markers on the fat: open and closed. The closed triangles indicate the location of the first use vessels. When the fat is placed on the dome, the closed triangle is aligned with the triangle marker on the dome shell. After all three vessels are used, the fat is rotated and the open triangles are aligned with the triangles on the dome. This ensures that the vessels are in the accurate location for the dissection exercises.
Triangles on skin 	The triangle markers on the skin are aligned with the triangles on the fat layer. These ensure that the puzzle piece lies directly over the vessel and that the tower markers align with the underlying magnets.
Cap placement notch 	The notch in the cap ensures that users place the cap in the correct orientation. Since the magnet divots are placed in the shape of a triangle, the cap has to be secured in a specific orientation for the magnet divots to align properly.

In the final HFP, the exercises existed as they would in the manufacturing phase. Final testing was performed in order to ensure that all specifications were correct and to build a specifications document, which was used to create final CGI and CAD files (Figure 9).

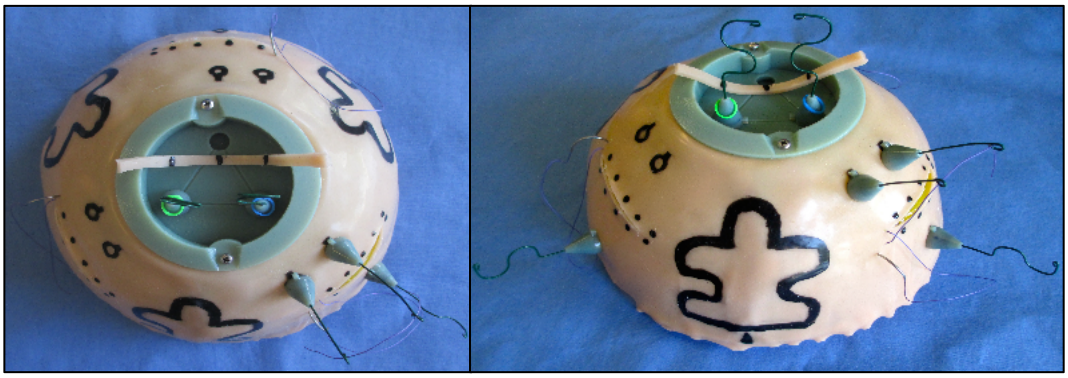


Figure 9. Final HFP

PRODUCTION

The final CGI, CAD, and specification document were sent to the manufacturing company and simulation companies to assist them in their development of physical and virtual domes (Figure 10).

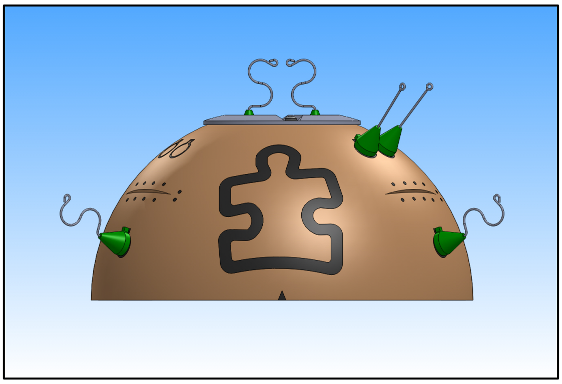


Figure 10. Final CGI

A local manufacturer, familiar with the materials used during prototype testing, used the dome and performed all of the exercises prior to beginning the process. This provided a first-hand experience of why certain material qualities were so important. The goals for this phase, in addition to mass production, were to maintain device integrity and minimize cost. Some of the materials used during prototyping were more expensive than what would be feasible for training centers. For example the $1,000 materials cost for the 3D printed dome was reduced to less than $25.

The simulation exercises of the FRS dome will be incorporated into two simulators: the da Vinci Skills Simulator (dVSS) and the Mimic dV-Trainer (Figure 11). Both systems contain the six FRS exercises, but vary in their software and hardware. The dVSS is a simulation system, which integrates with the actual console of the surgical system. This allows users to train using the exact hardware that they use when operating. The dV-Trainer is a standalone system that uses custom hardware and software. These simulations give the users experience performing the FRS exercises without requiring the use of the entire robotic surgical system. Generally, the systems are dedicated resources to the hospital surgical department and difficult to reserve for training purposes. The simulators also allow unlimited practice sessions without consuming the physical materials of the dome. The research team worked with each of the simulator companies to create and test multiple prototype versions of the exercise software. Our extensive experience with the real materials and our surgeons' experience with human surgery allowed us to critically evaluate the simulated behaviors of materials and the scoring methods. This feedback has led to significant improvements in the accuracy and usability of the simulators.

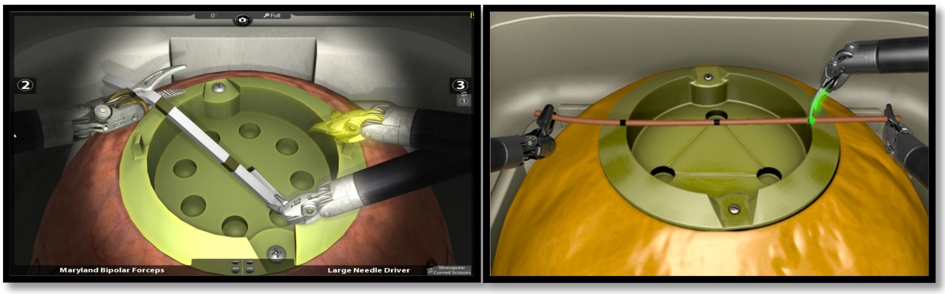


Figure 11. Mimic dV-Trainer and Simbionix's dVSS simulated dome exercises

Maintaining the simulated physical properties of the dome was paramount. Since the simulations may be used without proctors, the physical behaviors have a considerable impact on the scoring metrics and guidance that is given for improving performance. The research team evaluated the simulated exercise properties including elasticity of materials, flexibility of sutures, simulated gravity, and the effects of excess force on the virtual device to ensure that it behaved similar to the real dome. The real materials however were also limiting to some of the desired qualities, particularly in the vessel dissection exercise. The silicon-based materials act as insulators, preventing cauterization of the small vessel. Both simulators allow the user to apply energy for cauterization, as well as receiving a visual indication that the vessel is losing blood, prompting the user to manage the situation appropriately.

Some of the metrics also varied between the physical and simulated domes. While the physical dome is scored via expert video reviewing, the simulator can more objectively assess a user's performance. This allows the

simulated exercises to score some errors more accurately, such as instruments being out of view for a specific amount of time and over a specific distance.

The research team will include these simulations in a pilot study and provide the simulation companies further formative feedback on the usability of their systems, to mitigate complications that may occur during the larger multi-site validation study that will follow. This pilot study will also establish preliminary scoring benchmarks based on expert performance, which will be used to guide the multi-site validation study.

CONCLUSION

Over the course of two years, we created an easily integrated device, using low cost but high-quality materials. This paper outlines the steps of the FRS dome from idea conception to the development of physical and virtual devices. The goal of this paper is to share the evolution and process for others interested in training and assessment devices. Since the FRS dome specifications are open-source, this also serves as an important resource for potential producers.

We have taken away several lessons from our experimentation that made our process a success including having a multidisciplinary team, soliciting frequent feedback, using easily adaptable designs, testing on small models, and using commercial materials during prototyping. Our multidisciplinary team of surgeons and engineers allowed for a diverse perspective during the construction of the device. The design changed many times and it was beneficial to start off using basic models that accommodated the varying designs. It was advantageous to work with actual manufacturing materials once we developed a functional prototype to better envision the final product and allow a smoother transition to the manufacturing phase. We recommend testing materials on small models, which will help cut time and costs. Finally if possible, work closely with the manufacturing teams at an early stage of development, particularly when working with virtual models. This will help to flesh out details and encourage collaborative development earlier in the process.

The next step of this work is to conduct formal validation testing of the curriculum including the device and related simulations via a pilot and national multi-site validation study. The FRS dome features basic robotic surgical skill exercises, which are applicable to most specialties. This basic device is scalable and will be the foundation for the future, more specialized FRxS devices (e.g., the Fundamentals of Robotic Gynecologic Surgery (FRGS) and the Fundamentals of Robotic Urologic Surgery (FRUS)).

REFERENCES

Barbash, G. I., Friedman, B., Glied, S. A., & Steiner, C. A. (2014). Factors associated with adoption of robotic surgical technology in US hospitals and relationship to radical prostatectomy procedure volume. *Annals of Surgery*, *259*(1), 1–6. doi:10.1097/SLA.0b013e3182a5c8b8

Blavier, A., Gaudissart, Q., Cadière, G.-B., & Nyssen, A.-S. (2007). Comparison of learning curves and skill transfer between classical and robotic laparoscopy according to the viewing conditions: implications for training. *American Journal of Surgery*, *194*(1), 115–121. doi:10.1016/j.amjsurg.2006.10.014

Chitwood, W. R., Jr, Nifong, L. W., Chapman, W. H., Felger, J. E., Bailey, B. M., Ballint, T., … Albrecht, R. A. (2001). Robotic surgical training in an academic institution. *Annals of Surgery*, *234*(4), 475–484; discussion 484–486.

Chowriappa, A., Raza, S. J., Fazili, A., Field, E., Malito, C., Samarasekera, D., … Guru, K. A. (2014). Augmented Reality Based Skill Training for Robot-Assisted Urethro-Vesical Anastamosis: A Multi-Institutional Randomized Controlled Trial. *BJU International*. doi:10.1111/bju.12704

Geller, E. J., Schuler, K. M., & Boggess, J. F. (2011). Robotic surgical training program in gynecology: how to train residents and fellows. *Journal of Minimally Invasive Gynecology*, *18*(2), 224–229. doi:10.1016/j.jmig.2010.11.003

Grover, S., Tan, G. Y., Srivastava, A., Leung, R. A., & Tewari, A. K. (2010). Residency training program paradigms for teaching robotic surgical skills to urology residents. *Current Urology Reports*, *11*(2), 87–92. doi:10.1007/s11934-010-0093-9

Hubens, G., Coveliers, H., Balliu, L., Ruppert, M., & Vaneerdeweg, W. (2003). A performance study comparing manual and robotically assisted laparoscopic surgery using the da Vinci system. *Surgical Endoscopy*, *17*(10), 1595–1599. doi:10.1007/s00464-002-9248-1

Jarc, A. M., & Curet, M. (2014). Construct validity of nine new inanimate exercises for robotic surgeon training using a standardized setup. *Surgical Endoscopy*, *28*(2), 648–656. doi:10.1007/s00464-013-3224-9

Levy, J. (n.d.). Fundamentals of Robotic Surgery: Online curriculum. Retrieved from http://frs.casenetwork.com/lms/

Luca, F., Valvo, M., Ghezzi, T. L., Zuccaro, M., Cenciarelli, S., Trovato, C., … Biffi, R. (2013). Impact of robotic surgery on sexual and urinary functions after fully robotic nerve-sparing total mesorectal excision for rectal cancer. *Annals of Surgery*, *257*(4), 672–678. doi:10.1097/SLA.0b013e318269d03b

Maan, Z. N., Gibbins, N., Al-Jabri, T., & D'Souza, A. R. (2012). The use of robotics in otolaryngology-head and neck surgery: a systematic review. *American Journal of Otolaryngology*, *33*(1), 137–146. doi:10.1016/j.amjoto.2011.04.003

Patel, H. D., Mullins, J. K., Pierorazio, P. M., Jayram, G., Cohen, J. E., Matlaga, B. R., & Allaf, M. E. (2013). Trends in renal surgery: robotic technology is associated with increased use of partial nephrectomy. *The Journal of Urology*, *189*(4), 1229–1235. doi:10.1016/j.juro.2012.10.024

Serati, M., Bogani, G., Sorice, P., Braga, A., Torella, M., Salvatore, S., … Ghezzi, F. (2014). Robot-assisted Sacrocolpopexy for Pelvic Organ Prolapse: A Systematic Review and Meta-analysis of Comparative Studies. *European Urology*. doi:10.1016/j.eururo.2014.02.053

Smith R, Patel V, Satava R. (2013). Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. Int J Med Robot Comput Assist Surg MRCAS.

Smith, R., Patel, V. R., Chauhan, S., & Satava, R. (2013, February 20). *Fundamentals of Robotic Surgery: Outcomes Measures and Curriculum Development*. NextMed/ MMVR, San Diego, CA.

Zorn, K. C., Gautam, G., Shalhav, A. L., Clayman, R. V., Ahlering, T. E., Albala, D. M., … Members of the Society of Urologic Robotic Surgeons. (2009). Training, credentialing, proctoring and medicolegal risks of robotic urological surgery: recommendations of the society of urologic robotic surgeons. *The Journal of Urology*, *182*(3), 1126–1132. doi:10.1016/j.juro.2009.05.042

Zureikat, A. H., Moser, A. J., Boone, B. A., Bartlett, D. L., Zenati, M., & Zeh, H. J., 3rd. (2013). 250 robotic pancreatic resections: safety and feasibility. *Annals of Surgery*, *258*(4), 554–559; discussion 559–562. doi:10.1097/SLA.0b013e3182a4e87c

Developing Game-based Leadership Training for Robotic Surgeons

Roger Smith & Alyssa Tanaka
Florida Hospital Nicholson Center
Celebration, FL
Roger.Smith@flhosp.org
Alyssa.Tanaka@flhosp.org

Steve McIlwain & Brad Willson
ARA/Virtual Heroes
Raleigh, NC
Steve.McIlwain@virtualheroes.com
Bradley.Willson@virtualheroes.com

ABSTRACT

All surgeons must simultaneously perform as skilled practitioners and effective team leaders in the operating room. This is further complicated in robotic surgery because the surgeon is removed a short distance from the operating table and works from within a specialized cockpit. This separation creates a unique hurdle when a crisis arises that requires the surgeon to disengage from the immediate steps of the surgery to provide leadership and guidance with issues involving the team, the equipment, the room, or the patient.

To develop and test these skills we initially created a series of scenario-based videos with quizzes to evaluate surgeon understanding of these leadership responsibilities. Using these as a guide, we developed a game-based virtual environment containing the same information as the videos but in a 3D interactive space which is accessible through a web browser. This environment presents accessible and engaging scenarios that include a scoring mechanism which can assess the time to react to events, the actions that occur before and after a decision, and the correctness of the decision made. The tool can also present alternative or repetitive scenarios when the student does not take the correct action. This paper describes the development process and the interactions with the surgeons and operating room teams which drove the design and content of the virtual environment. The paper also describes the longer term plans to validate the content and introduction of the game to multiple surgical training sites around the country. Though the virtual environment uses a more interactive method for presenting leadership and team decision making information, we are interested in whether it is more effective than traditional didactic lectures, textual instructions, videos, and live role playing.

ABOUT THE AUTHORS

Roger Smith is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading-edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading research experiments. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *A CTO Thinks About Innovation*. He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

Alyssa D.S. Tanaka is a Systems Engineer at Florida Hospital Nicholson Center. Her research work focuses on robotic surgery simulation and effective surgeon training. Her current projects include rapid prototyping of surgical education devices, the validation of a robotic surgical curriculum and evaluation of robotic simulation systems. She is a Modeling and Simulation PhD student at the University of Central Florida and previously earned a M.S. in Modeling and Simulation, Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida. She holds a diploma in robotic surgery from the Department of Surgery, University of Nancy, France.

Steve McIlwain is a Senior Producer for the Virtual Heroes Division of Applied Research Associates, Inc. He has over 10 years of production experience in the 3D animation and interactive entertainment industries. Prior to joining Virtual Heroes, Steve worked at Walt Disney Feature Animation and Blizzard Entertainment. He specializes in production management, macro/micro scheduling, team building, and finance. He is passionate

about creating virtual worlds that educate, inform, and inspire. Steve holds a B.S. in Marketing and an M.B.A. from Azusa Pacific University

Bradley Willson is the Game Design Lead for the Virtual Heroes Division of Applied Research Associates, Inc. A nine-year veteran of the game industry, Brad began his career at Rockstar San Diego as a Development Support Supervisor, where he worked on various commercial game titles including the Midnight Club series, Table Tennis, and Red Dead Revolver. He joined Virtual Heroes in 2006, with the goal of incorporating his commercial game experience into games that had a true altruistic focus. At Virtual Heroes, Brad works to create, drive, and deliver the overall creative vision for the numerous serious games in development. Brad creates compelling software designs from conflicting viewpoints, communicates the designs to the development team, and translates the designs into detailed software mechanics, gameplay progression, and interface flow. Brad holds a B.S. in Wildlife Science from Purdue University.

Developing Game-based Leadership Training for Robotic Surgeons

Roger Smith & Alyssa Tanaka
Florida Hospital Nicholson Center
Celebration, FL
Roger.Smith@flhosp.org
Alyssa.Tanaka@flhosp.org

Steve McIlwain & Brad Willson
ARA/Virtual Heroes
Raleigh, NC
Steve.McIlwain@virtualheroes.com
Bradley.Willson@virtualheroes.com

INTRODUCTION

Surgery is a team sport requiring the coordinated activities of multiple healthcare professionals. This team assembles daily in different combinations for a few hours with the chief surgeon as the leader who is responsible for directing the surgical activity. Historically, the surgeon has been the most highly educated member of the team, the most socially respected, and the most dominant personality. This has created situations in which the surgeon manages the team as a dictator who does not listen to the experience-based concerns and educated input from the other members of the team. Organizations like the American College of Surgeons and the World Health Organization have responded to these issues by creating and propagating standard practices and training materials which promote cooperative participation by all members of the team and an open, inclusive attitude by the surgeon/leader of the team. The surgeon remains the primary person responsible for the outcome of the surgery, but is encouraged or required to solicit and apply the expertise of all members of the OR team.

Robotic surgery with the da Vinci surgical robot, the dominant device in the field, introduces additional challenges for keeping a team working together. Changes in the physical location and orientation of team members create one new hurdle in team cooperation. Figure 1a illustrates the positions of typical members of a surgical team for open and laparoscopic procedures. Everyone is physically clustered around the patient, within arm’s reach and easy speaking distance of one another. Direct eye-to-eye contact and communication is easy and directives to the team are difficult to confuse. By contrast, Figure 1b illustrates the positions of members of a robotic team. Most members remain at the bedside, but the surgeon has been separated from the encircled group. In order to operate the robot, the surgeon must remove himself from the bedside and take a position within a custom console to control the machine. This console pulls their physical actions, visual attention, and mental focus into an environment that is separate and unique from the rest of the team. This situation can potentially undermine the previous work that has been done to integrate the actions and expertise of the team within more traditional forms of surgery.

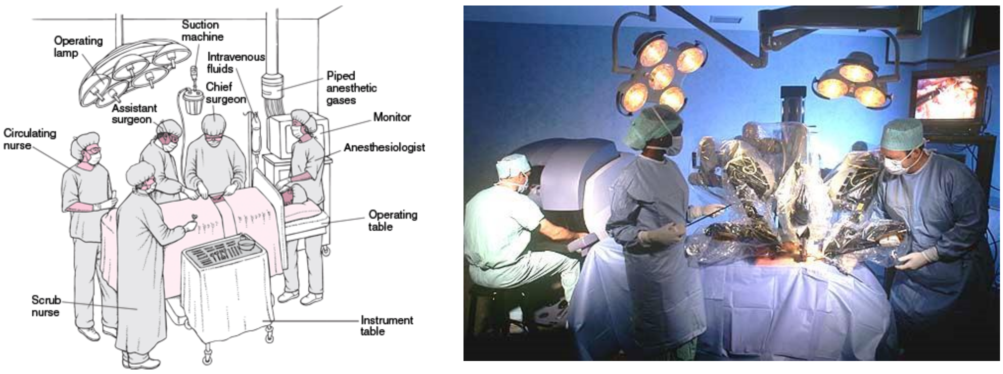


Figure 1. Traditional vs. Robotic OR Team Positions.

The manufacturer of the da Vinci robot has attempted to mitigate this separation by including a microphone and speakers in the head-space of the robotic console. So the words spoken by the surgeon are broadcast to the rest of the team from speakers attached to the bedside components of the robot. Similarly, a microphone on the bedside equipment captures the discussions of the surgical team and carries it to speakers in the surgeon’s console immediately next to the surgeon’s ears. External monitors around the bedside also display the picture of the internal surgery which the surgeon is seeing within the console. So all members share a common view of the inside of the patient and can talk to each other as if they remained around the bedside within arm’s reach of each other.

To teach and reinforce team management and leadership for surgeons there have previously been video instructions and role playing scripts that walk through each of the skills which have been identified as essential for surgical teams. The video recordings present previously enacted situations which can contain both correct and incorrect activities that the surgeon/student can be evaluated on through questionnaires following the video. But the situations do not require interactive participation by the surgeon. Live role playing events allow the surgeon and all of the actors to experience multiple variations on the situation and to explore unique ideas which emerge in real time. However, these are extremely difficult to coordinate and host. The working schedules of surgeons, circulating nurses, surgical technicians, and anesthesiologists are very different. Each profession is guided by different certifying boards, departmental management, educational requirements, and working hours. Arranging for live events within a hospital or at a professional conference can be nearly impossible with real professionals. At some educational conferences, these events have been organized using hired actors for the members of the team. These remain expensive and rare events. Though these methods have proven useful, some of their limitations may be overcome through a computerized, interactive, game-based learning environment.

This paper describes a project to create a surgeon leadership and team management virtual environment which could be used at a robotic surgeon's leisure. This environment can include more variations in activities than can be easily captured in videos and can provide some of the richness of live role-playing events, but without the expense and logistical hurdles.

This paper describes the process used to design, prototype, and field the virtual world application. The application is currently in final in-house testing and will be released for open community testing in the near future. After that it will become the basis for a validation trial focused on its educational effectiveness.

BACKGROUND

The robotic surgery team training virtual world (RoboTeamView) is the sixth product of a larger effort to create materials for the Fundamentals of Robotic Surgery (FRS) program, an authoritative, standardized curriculum for certifying the knowledge and skills of aspiring robotic surgeons (Smith, Patel, Satava, 2013).

The FRS program has leveraged the expertise of more than 50 of the leading robotic surgeons in the world as well as a number of educational and engineering professionals, to develop materials which surgical educators can use to bring new surgeons to a common, measurable, and professionally accepted level of proficiency prior to performing surgery on human patients (see Figure 2). These materials include:

a. Online Curriculum consisting of text, slides, photos, and videos for teaching the cognitive knowledge needed by robotic surgeons;
b. Psychomotor Skills Device which measures the tactile skills of a surgeon using the robot;
c. Team Training videos which convey material similar to that included in the RoboTeamView game;
d. Team Training Role Playing Script which can be acted by live role-players;
e. Intuitive Surgical da Vinci Skills Simulator (DVSS) exercises;
f. Mimic dV-Trainer simulator exercises;
g. Simbionix Robotix Mentor simulator exercises; and
h. Robotic Surgeon Team Training Virtual World (RoboTeamView) for teaching team skills to a surgeon who is training alone.

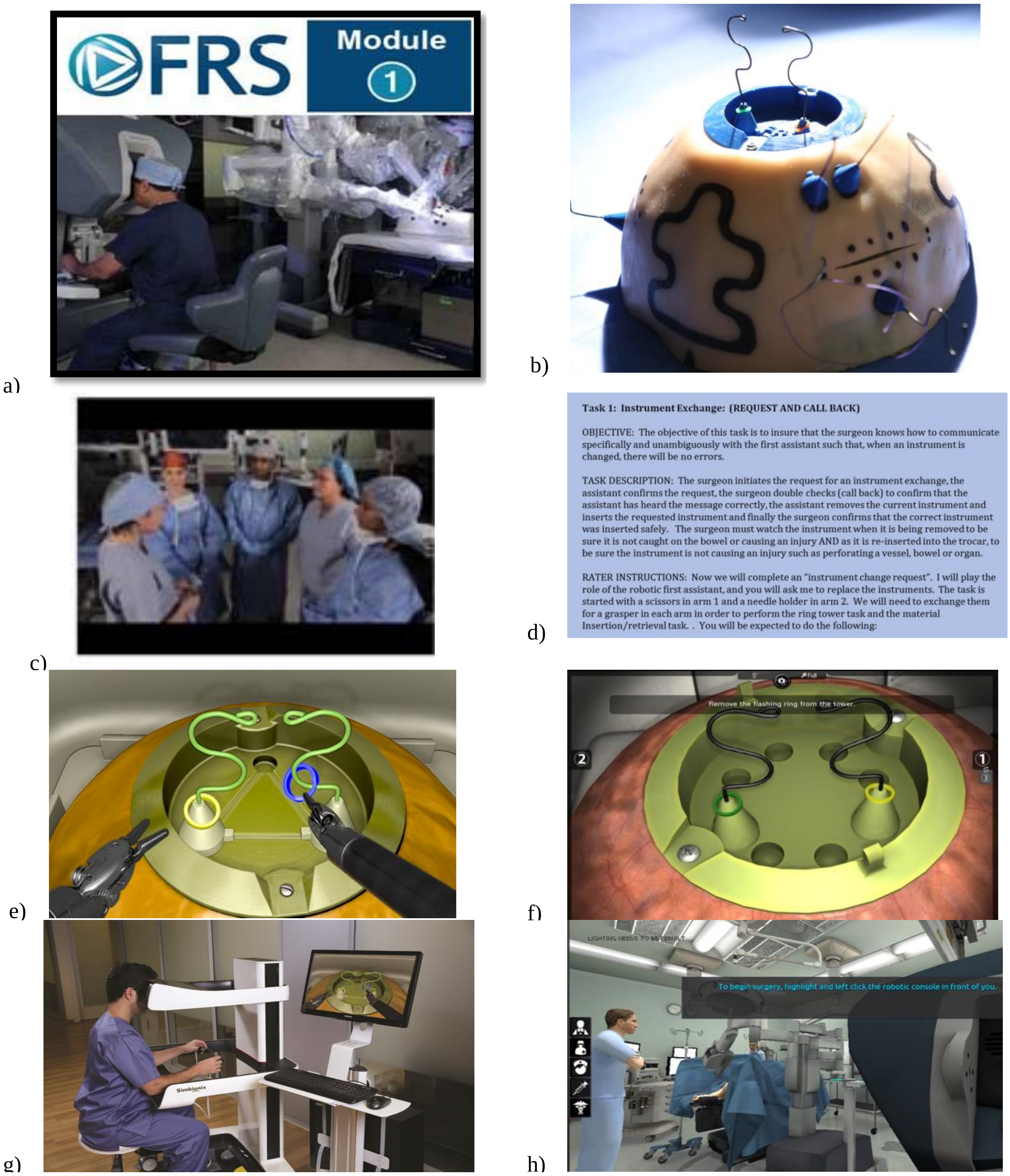


Figure 2. Fundamentals of Robotic Surgery Curriculum Products – (a) online curriculum, (b) psychomotor skills device, (c) team training videos, (d) role playing script, (e) Intuitive DVSS simulator, (f) Mimic dV-Trainer simulator, (g) Simbionix Robotix Mentor simulator, (h) RoboTeamView virtual world.

METHODOLOGY

This project used the ADDIE process for design and development of the learning application (Branch, 2010).

Analysis of Problem.

Surgeons with years of experience in bedside surgery (open and laparoscopic) described a sense of separation from the operating team when they began using the robot for procedures. In spite of the video and audio assistive tools which allow members of the team to communicate with each other, the physical separation and lack of direct line-of-sight to the team allowed the surgeon to immerse himself in a private world during a procedure. Effective communication with the team became something that required a higher level of conscious effort to maintain throughout a procedure. Surgeons needed to learn when to use the communication tools in the robot and when to disengage from the robot in order to handle situations which required more direct human-to-human contact (Hanly et al, 2006).

Analysis of Users.

There are two primary users of this virtual world. The first are attending surgeons, fellows, and residents who aspire to practice robotic surgery. The second are experienced robotic surgeons who require additional training in working effectively with a team. Both groups have limited time to focus on new curricula beyond their current work load. Both must learn independently in an environment that they access themselves. They do not have dedicated classrooms, equipment, instructors, and class hours as do traditional university students. In most cases, the student is expected to learn on their own time and without the collaboration of other members of the OR team.

Analysis of Environment.

The users typically possess extensive medical and surgical skills, but very limited computer skills. They are typically not proficient at installing new applications on computers, or they are using machines that are controlled by corporate IT restrictions which prohibit unauthorized applications. These characteristics led to a focus on a web-based application with a plug-in which auto installs if needed, and which can be approved for use across the corporate environment.

Design of Instruction.

Instruction is based on the widely used TeamSTEPPS curriculum (Safny et al, 2011; Thomas and Galla, 2013) and WHO checklists for surgery (WHO, June 2008). This material is then modified for application in a robotic OR environment. The exchanges with team members in this environment are largely prescribed and standardized to reduce miscommunication and the omission of important steps. The instructions for the game were based on prior work to create role-playing scripts for robotic OR team members.

Design of User Experience.

The primary instructional environment is a virtual robotic operating room which is populated with four avatars representing the other members of the team. The surgeon is either viewing a surgical field inside of a patient or the team around the operating table. In the former case, the surgeon interacts with a surgical video using menu selections at key decision points. In the latter, the surgeon queries an avatar for information and gives it instructions to be followed. The primary goal of the environment is to lead the surgeon through specific scenarios and assist them in understanding the correct actions that they should apply. This is primarily a learning environment and secondarily an assessment tool.

Development of Virtual Environment.

Virtual Heroes has previously created a number of healthcare virtual worlds which included digital assets that appear in this virtual world. The essential new asset which had to be created was a 3D model of the da Vinci surgical robot, a complex piece of machinery with many visible pieces. The robotic arms and hand controls need minimal animations for these team scenarios. More work was required for the multiple menu items necessary to present all of the decision actions of the team.

Development of Video Integration.

The project required the integration of 3D virtual world assets with prerecorded videos of the surgical field. These videos were drawn from the extensive video library of a leading robotic urologist and some videos were custom made during live surgeries. From these we were able to select segments of surgeries which corresponded to the lessons being taught in the virtual world. Synchronizing virtual actions with video events allowed us to avoid creating virtual representations of complex internal human anatomy and the manipulations of those models.

Development of Evaluation Criteria.

The scenarios provide multiple decision points at which a surgeon/student must select the correct response from a small list of options. The correct selection will lead to acceptance by the avatars and progression to the next step. An incorrect selection will cause the avatars to offer advice or to ask leading questions to guide the surgeon to a correct action. Performance evaluation is a summation of the correct and incorrect actions taken by the surgeon during each scenario. Benchmarking those scores will be part of a future validation process in which proficiency levels will be established based on the scores of expert and novice subjects.

Implementation of Training Program.

The training program will be implemented in multiple steps. Initially, the RoboTeamView will be made available to a small number of robotic surgeons who assisted with the development of the new curriculum. They will provide feedback during the early releases to assist in reprogramming or redesigning features of the application. The secondary release will be to a community of expert robotic surgeons who have contributed to the creation of previous FRS program materials. These experts are the conduits for sharing the application with aspiring robotic surgeons at multiple hospital systems and organizing a validation trial using surgeons, fellows, and residents. Finally, the application will be made publically available at no charge for access by anyone who is interested in using it for their own personal learning or as a tool within in an educational environment.

Evaluation of Effectiveness.

Acceptance of this material by instructors and institutions for education in robotic surgeon training is an encouraging and valuable achievement. But it does not constitute scientific evidence of validity as an effective teaching tool. This will be achieved via a multi-site validation trial of the tool with the goal of demonstrating that it is an equal or better method of teaching team leadership skills than the existing methods.

DEVELOPMENT

Data Acquisition

The development process began with the acquisition of knowledge and data. The game development team observed multiple procedures in the robotic operating room. They were able to watch and listen to all of the activities that occurred, and to see each member's role throughout a procedure. They also witnessed the transition of nursing support staff completing a shift or leaving for a break during a procedure. Following this exposure, robotic surgeons were interviewed, introduced to the product concept, and provided their guidance on how such a product could be structured for effective education. An analysis of the published literature of the use and availability of simulators or virtual worlds for robotic surgeons indicated that a leadership-focused tool for team communication skills had not previously been created (Kumar, Smith & Patel, 2015). Therefore, many of the educational design concepts of this project were being created for the first time.

The team reviewed existing curriculum in textual script and video recording formats. These were based on best practices which have been created by the TeamSTEPPS program and the World Health Organization for safe communications in the operating room. Together with the data collected from the surgeons, the team arrived at a small set of scenarios to be included in the virtual world, as listed in Table 1.

Table 1. Surgical Scenarios Created

S1.	Instrument Exchange (Request and Call Back)
S2.	Material Insertion & Retrieval (Request and Call Back)
S3.	Two-challenge Rule for a Safety Issue (CUS and SBAR)
S4.	Personnel Change (Handoff Responsibilities)
S5.	Check Back
S6.	Emergency Robotic Undocking Procedure
S7.	Pre-Brief (Checklist or Sign-in)
S8.	Post-procedure Debrief (Checklist or Sign-out)
S9.	Recoverable Robotic System Fault
S10.	Non-recoverable Robotic System Fault
S11.	Broken Instrument
S12.	Difficulty Removing/Reinserting an Instrument
S13.	Loss of Insufflation of Patient

The game calls for a combination of 3D computer graphic assets and live surgical videos. Through the cooperation of several surgeons the project received access to an extensive library of thousands of surgical videos. These videos are all usable for educational purposes through signed releases from the patients. As specific scenarios and 3D actions were developed, the team located an existing surgical video with actions which corresponded to the scenario. Using such a large video library made it possible to avoid either video recording a simulated surgery or attempting to create a realistic virtual representation of all of the surgical activities. In spite of the size of this library, it was necessary to custom record some actions during surgeries for this project. The current level of simulation technology is challenged to graphically model human tissue with manipulation, dissection, and blood flow. Some surgical VR simulators contain very realistic, but limited representations of surgery which require significant computer hardware to run. For this reason this project relies on video recording to represent actions in the surgical field, which comes with some inherent limitations to interactivity.

User Experience

Role Definition

Early discussions within the development team and with surgeons were focused on who would be the training audience for the tool. Since there are five members of the OR team who must learn to work together, should this tool provide a user interface and curriculum for each of these as potential trainees? Such a flexible tool seemed possible since the scenario is the same for each role, only requiring the removal of one script to allow a human user to play that part. However, since there were no previous tools of this type to use as guidance, solving such a multifaceted problem could lead to confusion and delays that would threaten the success of the project. Also, achieving acceptance of the tool from five different sets of professional and certifying organizations seemed to be a much larger problem. Therefore, the design focused only on training the surgeon, as was done with previous curriculum products. But, the virtual world and other training products may become the basis for variants that are targeted at the circulating nurse, first assistant, surgical technician, or anesthesiologist in the future. Since the game creates a single-user domain, there is no need for computer servers to coordinate the interactions of multiple players within the same scenario. A single scenario can be served to any number of users simultaneously, but each of these runs independently without the need for coordination between multiple players.

Dual Domains

During a procedure, the surgeon occupies two very different domains. One is as a member of the team that surrounds the operating table to address the patient from the outside. The other is a more private domain in which the surgeon is immersed within the internal anatomy of the patient with audio communication to the outside team (see Figure 3). In the scenarios which are to be represented (Table 1) it is most accurate for the surgeon to act within both of these domains, which requires creating a simulated environment of both. Previous training curriculum in video and script formats had presented the OR only from the external bedside view, while existing simulators provided only the internal view. This game is the first to include two very different domains in which the surgeon is learning. For some scenarios a surgeon remains immersed in the patient while responding to the team and giving direction. But for others, the surgeon needs to learn to disengage from the internal view in order to address a more important issue in the external OR. Learning which domain is most appropriate for the surgeon has become part of the training that is uniquely provided by this game.

(a) Virtual Operating Room

(b) Video Surgical Field

Figure 3. Virtual World Representation

Session Independence & Progression

When a surgeon proceeds through a scenario, their progress is stored on the local computer. This allows users to interrupt their progress through a scenario, but return to the same point when they pick-up the game at a future time. Information about progression is also exported to the Moodle Learning Management System (LMS) to provide scoring and evaluation of the players. When a surgeon chooses to terminate a scenario prior to completing it, the LMS has a record of progress that has been made. In future versions this information will make it possible for the surgeon to complete a scenario from multiple devices by loading past progress from the LMS. This capability is a potential extension should early users discover that it is an essential feature.

Security

Like most corporate environments, the hospital IT infrastructure is tightly controlled and monitored to protect again hostile external and internal actions. It also blocks certain private and social services which are not considered productive in a corporate environment. As a result, many ports and some data formats cannot be used by applications like this virtual world.

The application was designed for Windows 7 and 8 operating systems and the Internet Explorer v.7+ browser because these are the most common within the hospital. Virtual Heroes bases many of their custom projects on the Unreal engine licensed from Epic Games for simulation projects. This engine and the game content are configured as a one-time browser plug-in to eliminate issues with asking users to perform multiple heavy downloads and installations. As a plug-in, this process is largely automated upon first use of the application. However, corporate IT restrictions still verify that the plug-in is permitted within the controlled hospital infrastructure. Therefore, the plug-in was treated as a new application which had to be reviewed for security and stability issues before being allowed to enter a hospital computer.

Additionally, once installed, the plug-in communicates with the LMS via unique ports and data formats which had to be approved to traverse the hospital network (see Figure 4).

The application was originally developed and shared from a Virtual Heroes server, and was then tested on personal computers on an open commercial network. Once a basically functional version existed, a hosting site on the internet was created which required a fresh install away from the Virtual Heroes machines. This demonstrated that the application was portable enough to be hosted on a customer's servers as opposed to the developer's servers. Finally, the hospital IT department created a hosting site within the hospital infrastructure, approved the plug-in on hospital computers, and opened the necessary communication ports for the application. The goal is to host the application on a site which can be accessed by surgeons both inside and outside of the hospital infrastructure. Robotic surgeons who are not employees of Florida Hospital will access the external site.

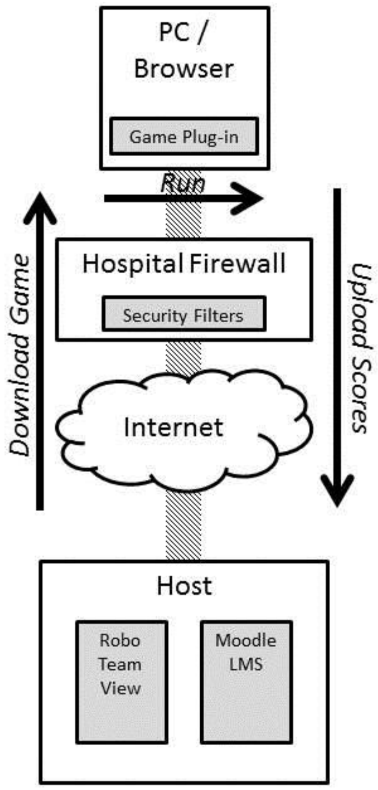


Figure 4. RoboTeamView System and Network Architecture

User Evaluation

The performance of the surgeon is evaluated as they interact with the scenarios and the dynamic avatars in the game. The application provides very direct guidance regarding the steps that are expected. The intention was for the game to be more of an educational environment than an assessment tool. The design allows surgeons to work through the scenarios without a human instructor and to learn the necessary information for performing as a team leader. There are numerous opportunities for a surgeon to make decisions, each of which is captured in the LMS to provide some measure of their performance. But, an attentive surgeon can learn the correct responses from the avatars without having to consult a human instructor. Therefore, the measurements of performance are actually a measure of the surgeon's ability to learn and adapt to the guidance of the avatars in the game.

Each surgeon logs into the system to create a record of on-going performance in the LMS. The Moodle LMS also provides login for an instructor who can access all student performance data. This allows a hospital, college, or education center to track the performance of their people and to insist on a specific level of mastery in association with credentialing, risk management, and educational progression.

VALIDATION AND DEPLOYMENT

The virtual world application has been completed and is being evaluated by experienced robotic surgeons and teaching faculty at Florida Hospital. The feedback from these professionals will be incorporated into the application before releasing it to a larger audience for independent and objective validation trials. The FRS project has developed research relationships with a number of leading medical institutes around the world. These have participated in the validation of previous FRS products and have shown their ability to organize and conduct these types of trials. The sites listed in Table 2, as well as others who have shown interest in the materials, will be invited to access this application and participate in a multi-site validation trial.

Following these trials, the revised application will be made available on the TrainRobotic.com web site for aspiring robotic surgeons, instructors, and medical training facilities to use as a curriculum for training robotic surgeons in their leadership responsibilities within the OR. Users of the application will be able to track student performance via the linked LMS.

Table 2. Robotic Surgery Curriculum Validation Site List

Florida Hospital Nicholson Center, Orlando FL	Lahey Health and Medical Center, Boston MA
University of Athens Medical School, Greece	Hartford Hospital, Boston MA
Imperial College, London UK	Louisiana State University School of Medicine, New Orleans
EndoCAS, Pisa Italy	Madigan Army Medical Center, Seattle WA
Baylor University Medical Center, Dallas TX	University of South Florida Health CAMLS, Tampa FL
Carolinas Healthcare System, Charlotte NC	Methodist Medical Center MITIE, Houston TX
Lehigh Valley Health Network, Allentown PA	University of Pennsylvania Medical Center, Philadelphia
Duke University Medical Center, Raleigh NC	

CONCLUSIONS

The primary goal of this project was to determine whether an effective leadership training application could be created for robotic surgeons who must learn to lead a team in the OR while performing surgery. The bulk of the efforts went into identifying which scenarios should be represented and how the information should be structured to create an effective training tool. The resulting product demonstrates that such an application can be created and that it satisfies potential users. As of this writing, the tool has not been used to train surgeons, fellows, or residents in OR team leadership. Neither has a validation trial been conducted to compare the effectiveness of this method against existing methods, e.g. didactic lectures, textual instructions, video recorded cases, and live role playing events. The next step is to conduct such a validation trial to determine whether the application is effective at teaching these skills to robotic surgeons. The results of these experiments and educational experiences are potential topics for future publications.

Questions that remain outstanding include:

- Will experts and instructors incorporate the application into their curriculum?
- Do surgeons who use the application actually have better patient outcomes?
- Is the application better than or equal to existing methods of teaching these skills?
- Is the product sustainable over a period of years, both financially and as educational content?

REFERENCES

Branch RM. (2010). *Instructional Design: The ADDIE Approach*. New York, NY: Springer.

Hanly EJ, Miller BE, Kumar R, Hasser CJ, Coste-Maniere E, Talamini MA, Aurora AA, Schenkman NS, Marohn MR. (Oct 2006). Mentoring console improves collaboration and teaching in surgical robotics. *Journal of Laparoendoscopic Advanced Surgical Techniques*.16(5):445-51.

Kumar AA, Smith RD, Patel VR. (March 2015). Current status of robotic simulators in acquisition of robotic surgical skills. *Current Opinions in Urology*. 25(2):168-74.

Sanfey H, McDowell C, Meier AH, Dunnington GL. (Feb 2011). Team training for surgical trainees. *Surgeon*. 9 Suppl 1:S32-4.

Smith RD, Patel VR, & Satava RM. (Sept 2014). Fundamentals of robotic surgery: a course of basic robotic surgery skills based upon a 14-society consensus template of outcomes measures and curriculum development. *The International Journal of Medical Robotics and Computer Assisted Surgery*. 10(3):379-84.

Thomas L, Galla C. (May 2013). Building a culture of safety through team training and engagement. *BMJ Quality and Safety*. 22(5):425-34.

World Health Organization. (June 2008). *Implementation Manuel Surgical Safety Checklist*. Retrieved February 1, 2015 from http://www.who.int/patientsafety/safesurgery/ss_checklist/en/

Gamers Today, Surgeons Tomorrow?

Alyssa Tanaka, M.S., Courtney Graddy, M.S., Roger Smith, Ph.D.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org, courtney.graddy@flhosp.org, roger.smith@flhosp.org

Manuela Perez, M.D., Ph.D.
Nancy University Hospital
Nancy, FR
m.perez@nancy-chu.fr

ABSTRACT

Faced with an age of reliance on technology and innovative advances, surgeons are using cutting-edge robotic systems to perform complex procedures and virtual reality simulators for specialized skill training. The virtual environment and controllers in surgical simulators are reminiscent of those in videogames. So, can playing video games develop skills similar to those used in robotic surgery?

This paper compares the performance of video gamers, medical students, and "lay people" to expert robotic surgeons on a robotic surgery simulator. Participants recruited from the UCF College of Medicine, UCF FIEA, and Florida Hospital completed a demographic questionnaire. The subjects then performed three computer-based perceptual tests and participated in two warm-up tasks on the Mimic dV-Trainer to familiarize themselves with the system. The experiment then measured their performance over eight trials of two core simulated exercises. After completing these trials, participants completed a post-questionnaire about their experience.

Analysis of the data did not verify differences between the groups for the perceptual tests except for the time to complete scores in the Flanker and subsidizing tasks, in which expert surgeons took significantly longer than other groups. Significant differences were found between the groups for the first and eighth trials of the simulated exercises, with surgeons performing better than other groups. All groups improved significantly from trial one to trial eight, with surgeons performing better than all groups. Gaming console type positively correlated with Overall Score in the Ring & Rail exercise, as well as Time and Economy of Motion in the suturing exercise. No other correlations were found.

The results are in contrast with prior literature on video game experience in laparoscopic surgery, suggesting that gaming abilities do not translate to all surgical modalities. Future research is necessary to further examine the impact alternative skillsets may have on surgical skills.

ABOUT THE AUTHORS

Alyssa D.S. Tanaka, M.S. is a Research Scientist at Florida Hospital Nicholson Center. Her research work focuses on robotic surgery simulation and effective surgeon training. Her current projects include rapid prototyping of surgical education devices, the validation of a robotic surgical curriculum and evaluation of robotic simulation systems. She is a Modeling and Simulation PhD student at the University of Central Florida and previously earned a M.S. in Modeling and Simulation, Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida. She holds a diploma in robotic surgery from the Department of Surgery, University of Nancy, France.

Courtney Graddy, MHA is a Human Studies Research Coordinator at the Celebration Health Research Institute where she manages projects aimed at improving patient health outcomes, employee health, process improvement and simulation research. Her current projects focus on integrating technology into standard of care and evaluating its effects on patient health and patient satisfaction, as well as evaluating teaching modalities used to train surgeons. Her career began at the North Florida South Georgia Veterans Health

System where she aided in the development of employee education materials and program planning and evaluation with the Geriatric Research Education and Clinical Center. She holds a Bachelors of Science in Health Education from the University of Florida and a Masters of Health Administration from the University of South Florida.

Manuela Perez, M.D., Ph.D. is a practicing General Surgeon at the University Hospital of Nancy-France, where she also serves as an Assistant Professor in General Surgery and Anatomy. Dr. Perez has been practicing medicine for 14 years and graduated with her PhD in Robotic Surgery, with a thesis entitled "Telesurgery: From Training to Implementation." Currently, she is working as a Research Fellow at the Florida Hospital Nicholson Center and working under a grant from the Department of Defense researching various aspects of Telesurgery.

Roger Smith, Ph.D. is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading-edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading research experiments. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *A CTO Thinks About Innovation*. He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

Gamers Today, Surgeons Tomorrow?

Alyssa Tanaka, M.S., Courtney Graddy, M.S., Roger Smith, Ph.D.
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.tanaka@flhosp.org, courtney.graddy@flhosp.org, roger.smith@flhosp.org

Manuela Perez, M.D., Ph.D.
Nancy University Hospital
Nancy, FR
m.perez@nancy-chu.fr

INTRODUCTION

Surgery is generally described as fitting into one of two modalities—open and minimally invasive, the latter of which includes laparoscopic and robotic-assisted (i.e. robotic surgery) procedures. Robotic surgery, the most recent iteration of laparoscopy, typically implies that the surgeon's movements are facilitated through a computer driven system to manipulate surgical tools. This field evolved from the prospect of surgeons performing life saving procedures on soldiers in combat zones from remote locations anywhere in the world, an application referred to as telesurgery.

This concept has not completely come to fruition, however the fundamental research resulted in the commercial the daVinci Surgical System that is now used to perform everyday procedures in urologic, gynecologic, ENT, and general surgery specialties called the daVinci Surgical System (Barbash, Friedman, Glied, & Steiner, 2014; Serati et al., 2014; Maan, Gibbins, Al-Jabri, & D'Souza, 2012; Luca et al., 2013; Zureikat et al., 2013). The surgeon manipulates controllers at the surgeon console to manage up to four robotic arms, including a camera, attached to a separate patient cart. The camera provides true stereoscopic vision to the surgeon, facilitating a synthetic tactile sensation and depth perception. Attached to the other robotic arms are various instruments, which move in a similar manner as the surgeon's hands (Figure 1).

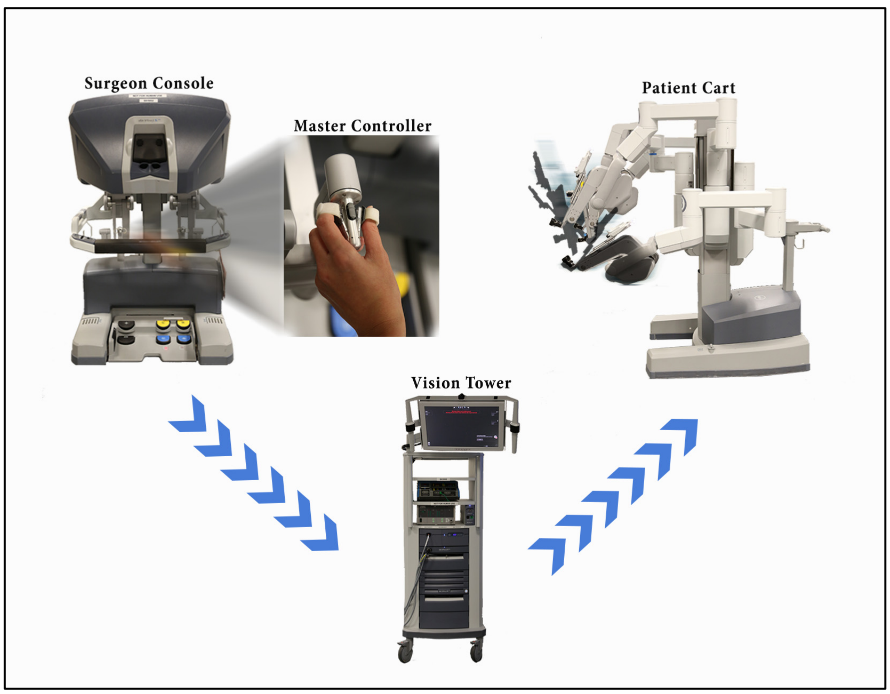


Figure 1. The daVinci System

While this system integrates robotics into medicine in a way that may seem more science fiction than reality, society is actually connecting with technology in unforeseen ways. Traditional surgical skills are being transcended by cutting-edge technologies, which require surgeons to possess distinct skill sets from those of the past and which overcome a learning curve to acquire the technical (i.e. psychomotor) skills associated with using the daVinci system. Efforts have focused on developing specialized curricula for the training of such skills (e.g. the Fundamentals of Robotic Surgery and Robotic Training Network), but can learning curves be reduced to facilitate a faster acquisition of skills in surgical trainees?

Previous research has established that trainees with video game experience demonstrate increased abilities on basic laparoscopic skill trainings (Rosenthal et al., 2011; Grantcharov, Bardram, Funch-Jensen & Rosenberg, 2003; Rosser et al., 2007). Also, video games have proven to be valuable training tools for basic laparoscopic skills (Rosser, Gentile, Hanigan, & Danner et al., 2012; Badurdeen et al., 2010; Ju, Chang, Buckley, & Wang, 2012; Bokhari et al., 2010; Schlickum, Hedman, Enochsson, Kjellin, & Fellander- Tsai, 2009; Giannotti et al., 2013). Certain genres of video games have established effects on perceptual skills similar to those required by robotic surgeons, yet few have attempted to make a connection between video game experience and robotic surgical skills (Green & Bavelier, 2012; Green & Bavelier, 2007; Chien et al., 2013; Harper et al., 2007).

Thus, this research aims to examine the performance of experienced video gamers while using a robotic surgery simulator, and compare the performance of this population with experienced robotic surgeons, medical students, and laypeople. The purpose is to determine the effect that video game usage may have on the perceptual abilities that are used for robotic surgery. Contrary to previous research that used surgical trainees with minimal gaming experience, this research aimed to utilize subjects with high levels of gaming experience and compare their abilities to subjects with different levels of expertise. This study also looks at the groups' ability to acquire basic surgical skills using the simulator.

METHODS

Recruitment

Participants in this study included video game experts (VGEs), expert robotic surgeons, medical students, and "laypeople" (i.e. individuals without formal medical education or extensive gaming experience). VGEs were recruited from a local university offering degrees specializing in game design and development (i.e. Florida Interactive and Entertainment Academy [FIEA]). Potential VGE subjects were required to be enrolled in a game design program and self-report daily videogame play of at least two hours per day, five days per week. Expert robotic surgeons were recruited from Florida Hospital, Florida Hospital Nicholson Center training courses, and at relevant surgical conferences. These individuals were practicing physicians and self-report performing at least 100 robotic surgical procedures, of which he or she performed at least 50% of the procedure on the surgical console. Medical students were recruited from the University of Central Florida College of Medicine (UCF CoM) and laypeople were recruited from all data collection sites. Potential subjects were excluded from the study in the case of having experience in more than one participant category (e.g. a medical student or expert robotic surgeon who engages in regular gameplay of more than two hours per week).

Materials

All subjects completed a pre-questionnaire, which gathered demographic information (e.g., age, gender, handedness, hours of weekly gameplay, number of robotic cases). The participants then performed three computer-based perceptual tests: a Flanker compatibility task, a subsidizing task, and a Multiple Object Tracking (MOT) test. The Flanker compatibility test requires the participant to indicate the orientation of a single arrow in the center of a group of several other arrows. The arrows are randomly generated to all face the same orientation (congruent) or face the opposite direction of the target arrow in the center (incongruent). This tests attentional capacity by requiring the subject to focus solely on the relevant arrow and ignoring other stimuli. The subsidizing task also assesses attentional capacity by requiring subjects to identify the number of dots that appear on the screen by pressing the associated number key. In the MOT

task, users must track specific objects while they move across the screen with other identical objects, which assesses visual attention (Figure 2).

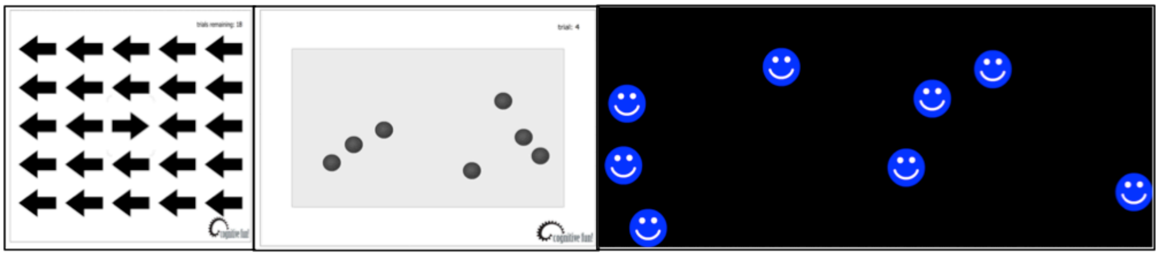


Figure 2. Examples of the Flanker, subsidizing, and MOT tasks

Participants then performed two warm-up exercises on the Mimic dV-Trainer, Pick & Place and Basic Camera Targeting, to familiarize themselves with the system and system controls. All subjects then performed eight trials of two core exercises to test various basic skills (Table 1). Ring & Rail 1 and Suture Sponge 1 will serve as the primary exercises for data collection. After completing all exercises on the dV-Trainer, specific metrics are shown to the participants: Overall Score, Economy of Motion, Time to Complete, Excessive Instrument Force, Instruments Out of View, and Master Workspace Range. These primary metrics are exported for each exercise and used with other metrics to form the scoring system.

Table 1. dV-Trainer exercise descriptions

Exercise	Purpose	Objective	Skills Trained
Warm-up Exercises			
Pick & Place	Introduction to using stereo vision and EndoWrist instruments for picking up and placing objects.	Place colored objects in matching colored containers.	Endowrist Manipulation
Basic Camera Targeting	Learn to accurately position the camera while working in a large workspace while practicing to keep the instruments in view and developing stereo depth acuity.	Manipulate the camera to position light blue sphere camera targets in the center of your screen's dark blue crosshairs.	Camera Control
Core Exercises			
Ring & Rail 1	Coordinate control of an object's position and orientation along a trajectory using the EndoWrist instruments	Pick up a ring and guide the ring along a curved rail	Endowrist manipulation, Camera Control
Basic Suture Sponge	Improve dexterity and accuracy when driving a needle through a deformable object.	Insert and extract a needle through several targets on the edge of a sponge with random variations in their positions.	Endowrist manipulation, Camera Control, Needle Control, Needle Driving

After completing all trials, participants completed a post-questionnaire regarding their experience with the system (Figure 3).

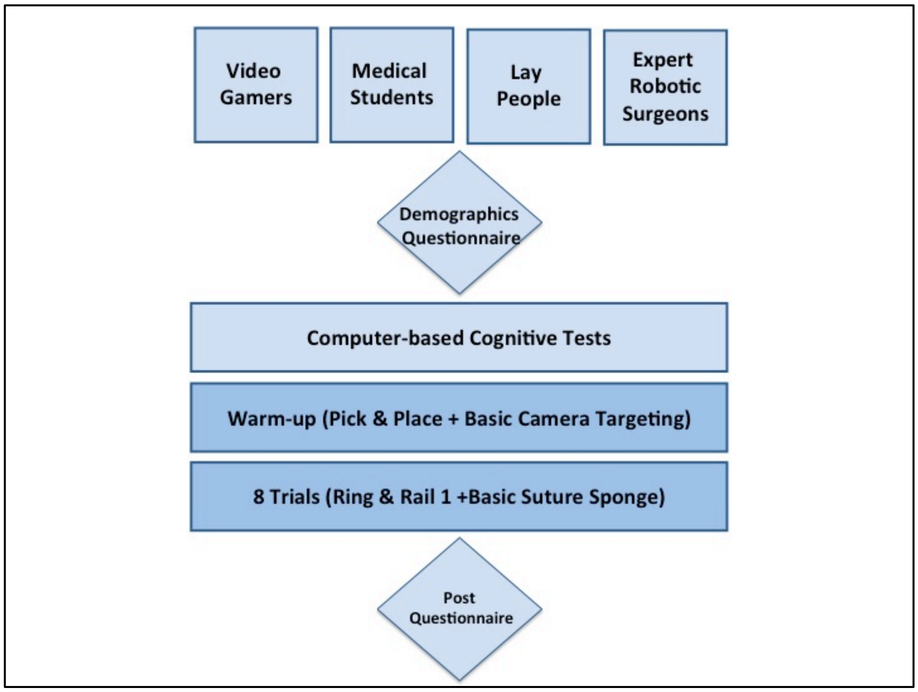


Figure 3. Order of study procedures

RESULTS

Demographics

Table 2 shows descriptive characteristics of the participants. Gamers indicated playing on average 11.71 hours of video games per week and having 17.85 years of gaming experience. On average, expert robotic surgeons performed 503 total robotic cases and 127 cases per year. While none of the expert surgeons reported currently playing video games, 29% indicated playing video games in the past. Thirty-three percent of lay people also indicated playing video games in the past.

Table 2. Descriptive statistics

Descriptive Statistics				
	Gamers	Medical Students	Laypeople	Experts
n=	40	24	42	7
Age	25.38	25.63	29.45	42
Male	77.5%	70.83%	52.38%	71.43%
Female	22.5%	29.17%	47.62%	28.57%
Right Handed	87.50%	95.83%	83.33%	100%
Left Handed	12.50%	4.17%	16.67%	0%

Cognitive Tests

For the Flanker and the subsidizing tasks, an ANOVA was performed to compare the four groups in terms of percent of correct responses and average response time (ms) for incongruent and congruent arrows. No statistical differences were found for the percent correct for the Flanker test, however completion times for the congruent and incongruent representations were significantly different between the groups (Congruent $p<0.005$; Incongruent $p=0.007$). Expert robotic surgeons took longer in both instances to perform the tasks (Table 3).

Table 3. Analysis of cognitive tests

Descriptives

	Flanker					***Subsidizing***				
	Percent Correct	*Std. Dev*	*Congr. Time*	*Std. Dev*	*Time Incongr.*	*Std. Dev*	*Percent Correct*	*Std. Dev*	*Time*	*Std. Dev*
Gamers	97.37	3.63	438.72	54.24	487.60	60.21	80.97	11.25	921.40	116.87
Medical Students	97.62	4.90	410.14	45.92	465.85	64.70	74.36	14.34	957.99	148.45
Lay people	97.98	3.32	469.81	88.86	525.70	93.84	74.21	14.95	991.94	138.00
Experts	98.57	2.44	525.12	147.32	554.65	95.24	71.93	13.40	1133.64	84.22

ANOVA

		df	**F**	**Sig**
Flanker	Percent Correct	3, 107	0.303	.823
	Congruent Time	3, 106	5.358	.002
	Incongruent Time	3, 107	4.285	.007
Subsidizing	Percent	3, 109	2.310	.081
	Time	3, 109	5.980	.001

No significant differences were found for the percent correct on the subsidizing task for any groups using a Kruskal-Wallis test. Similarly to the Flanker test, completion times were significantly different between the groups ($p=0.001$), with expert surgeons performing slower than the other groups. The MOT test was analyzed using a non-parametric test to compare the number of correct responses. No significant differences were found for any groups for the MOT test.

The cognitive scores were also analyzed in terms of certain demographic responses to determine if an association exists between the demographic characteristics and the cognitive test scores. A Pearson correlation coefficient was calculated. The characteristic of age positively correlated with the Flanker Time ($p=0.008$) and Flanker Incongruent Time ($p<0.005$). Age negatively correlated with the hours of weekly video game play ($p=0.010$). Age was also negatively correlated with the number of correct responses in the normal level of difficulty MOT task ($p<0.001$).

Simulator Scores

The simulator scores were analyzed in terms of three performance metrics for both simulated exercises: Overall Score, Economy of Motion, and Time to Complete. Overall Score is a composite score comprised of multiple performance metrics, including Economy of Motion and Time to Complete. Economy of motion is the total distance that the instrument tips moved and is measured in centimeters. Time to Complete is the total number of seconds required by the user to perform the exercise.

An ANOVA was used to determine if differences existed between the groups for the first (i.e. Trial 1) and the last (i.e. Trial 8) of the Ring & Rail 1 and Suture Sponge for the performance metrics. The groups performed significantly different for the performance metrics for trial 1 in both exercises except for the Overall Score of Ring & Rail. Using a Least Significant Difference Test, experts performed significantly better than other groups for the metrics. Similar results were found for trial 8 of both exercises. Experts again performed significantly better than all groups in trial 8 for both exercises scores all metrics (Table 4).

Table 4. Analysis of simulator scores for Trial 1 and Trial 8

ANOVA							
Trial 1				**Trial 8**			
Ring & Rail 1							
	df	***F***	***Sig.***		***df***	***F***	***Sig.***
Overall Score	3, 112	2.251	0.086	**Overall Score**	3, 112	1.369	0.256
Economy of Motion	3, 112	2.795	< 0.05	**Economy of Motion**	3, 112	6.314	< 0.005
Time	3, 111	5.050	< 0.005	**Time**	3, 112	5.278	< 0.005
Suture Sponge							
Overall Score	3, 112	8.948	< 0.001	**Overall Score**	3, 112	4.316	< 0.05
Economy of Motion	3, 112	5.175	< 0.005	**Economy of Motion**	3, 112	5.518	< 0.005
Time	3, 112	9.244	< 0.001	**Time**	3, 112	8.383	< 0.001

The simulator performances were also analyzed using an ANOVA to determine if differences exist between the groups in terms of the change in performance from trial 1 to trial 8 for both exercises separately (Table 5). A difference existed in the average Overall Score and Economy of Motion metrics from trial 1 to trial 8 for all groups in the Ring & Rail 1 exercise (Overall Score $p<0.001$; Economy of Motion $p<0.001$). Experts were found to be significantly different from the other groups for both metrics (Overall Score $p=0.045$; Economy of Motion $p=0.002$). A significant interaction was found between the trials and the groups for the Time metrics ($p=0.006$). The main effects of the trials were not examined due to this interaction.

Table 5. Analysis of change in simulator scores from trial 1 to trial 8

ANOVA						
	Ring & Rail 1			**Suture Sponge**		
	Overall Score	***Economy of Motion***	***Time to Complete***	***Overall Score***	***Economy of Motion***	***Time to Complete***
df	3, 109	3, 109	3, 108	3, 109	3, 109	3, 109
F	2.772	5.468	5.583	8.520	6.887	12.641
Sig	.045	.002	.001	< .001	< .001	< .001

A difference existed in the average Overall Score and the Economy of Motion metrics from trial 1 to trial 8 for all groups in the Suture Sponge exercise (Overall score $p<0.001$; Economy of Motion $p<0.001$). Experts were also found to be significantly different from the other groups for both metrics (Overall Score $p<0.001$; Economy of Motion $p<0.001$). A significant interaction was found between the trials and the groups for the Time metric ($p=0.011$). The main effects of the trials were not examined due to this interaction. The average of each metric across the eight trials for each exercise can be seen in Figure 4 and Figure 5.

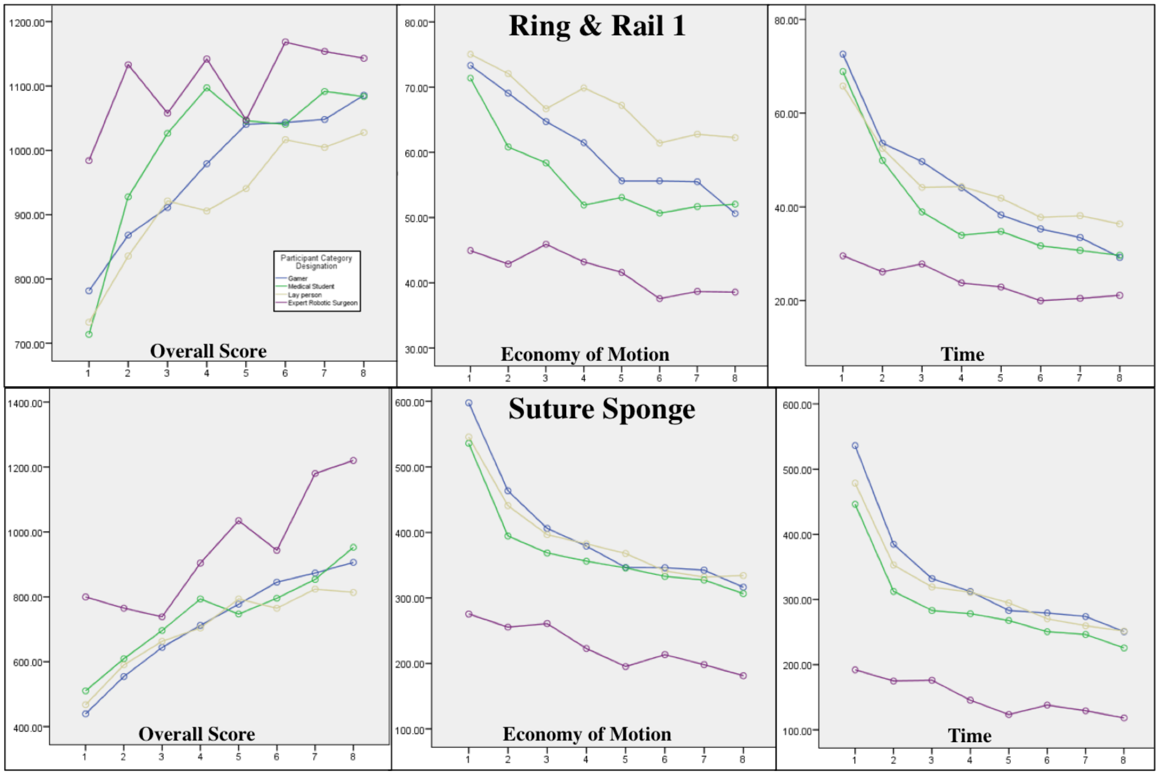


Figure 4. Average scores for groups across eight trials

An analysis was conducted to determine if an association existed between the perceptual test scores and the simulator metrics for the two exercises. The Flanker scores for the percent of correct responses negatively correlated with time to complete for the Ring and Rail 1 exercise (p=0.006). This suggests that as the correct response percentage increased the time taken to complete the exercise decreased. No other Ring & Rail 1 metrics correlated with the perceptual tests. No associations were found between the Suture Sponge scores and the perceptual test scores. The subsidizing and MOT task scores were not significantly correlated with any metric values for Ring and Rail 1 or Suture Sponge.

Video Games

The video game experience of the subjects was also analyzed to determine if certain aspects of video game play were associated with simulation scores. For this analysis the type of game and console played by the subjects was used. The game type ranged from not using videogames, playing slow-paced strategy games (e.g. puzzle games), playing both types of games, or playing fast-paced action games (e.g. first person shooters). The console type ranged from not playing video games, using a controller with minimal hand movement (e.g. Playstation4), using all controller types, or using a controller that may require larger movements (e.g. Wii).

No significant correlations were found between the type of video game or console played and the performance metrics for either exercise for trial 1. A significant positive correlation for Overall Score and the type of console was found for trial 8 of Ring and Rail 1 (p=0.049). This association suggests that as the movement to control the game increased, the Overall Score increased. A significant positive correlation was found between the type of console and Economy of Motion and Time for trial 8 of Suture Sponge (Economy of Motion p=0.044; Time to Complete p=0.002). This suggests that as the movement to control the video game increased, the time to complete and the distance traveled by the instrument tips increased (i.e. slower and less efficient with movements).

DISCUSSION

The assumption that video gamers will perform better than others using a virtual reality robotic surgery simulator is very common. The manipulation of the hand controls and the users interaction with the synthetic environment seem comparable to that of a video game. Contrary to these similarities and prior literature in laparoscopy, video gamers in this study did not perform better than other groups including the "Average Joe" in a robotic surgery simulator. The results did suggest that subjects who use higher movement game controllers (i.e. Ninetendo Wii) scored higher in the Ring & Rail 1 exercise. However, those individuals also took longer and were less efficient with their movements in the Suturing exercise.

The results from this study align with the few studies that have examined the impact of video game play on robotic surgical skills. Chien et al. (2013) found that in comparison to a group using task specific virtual reality training, a control group using video game training did not perform as well on an actual task using the surgical robot. The authors also found that using a video game to train actually had a negative impact on the post-training performance. Harper et al. (2007) found that video game players tied significantly fewer knots using the surgical robot and also suggest that video games may have a negative impact on surgical skills.

Why does prior video game experience impact basic laparoscopic skills, but not robotic? Differences may be contributed to the distinctness of the systems that the users are interacting. The skills developed in two-dimension video games may transfer more appropriately to laparoscopic surgery, which uses a two-dimensional screen, as opposed to the three-dimensional view in robotics. Laparoscopy involves contrasting movements to the primarily fine motor movements of robotic surgery and it is possible that gamers are more inclined with the manual dexterity associated with laparoscopy.

While this study was unable to validate enhanced abilities of video gamers in robotic surgery, the results demonstrated that the effect video game play has on surgical skills is nuanced by the surgical technique. In a technologically dependent society where video games have become an integral past time, this analysis of skills will likely become more valuable as other fields leverage the gaming generation's experience into training. The findings can be generalized to domains outside of medicine utilizing robotic and computer-controlled systems (e.g. unmanned vehicle operation), speaking to the scope of the gamers' abilities and pointing to the capacity within these systems.

Future research should examine the impact alternative skillsets may have on a user's abilities in a robotic surgery system (e.g. playing sports). The gamers in this study did not perform significantly better than lay people, which may imply that other factors or hobbies contributed to the performance. Only one surgical robot currently exists, however others have realized the technological advances and future iterations of surgical robotic systems are imminent. As these new technologies enter the market, it will be critical to evaluate how these skillsets may be valuable to the field of robotic surgery.

REFERENCES

Badurdeen, S., Abdul-Samad, O., Story, G., Wilson, C., Down, S., & Harris, A. (2010). Nintendo Wii video-gaming ability predicts laparoscopic skill. Surgical endoscopy, 24(8), 1824-1828.

Bokhari, R., Bollman-McGregor, J., Kahol, K., Smith, M., Feinstein, A., & Ferrara, J. (2010)

Barbash, G. I., Friedman, B., Glied, S. A., & Steiner, C. A. (2014). Factors associated with adoption of robotic surgical technology in US hospitals and relationship to radical prostatectomy procedure volume. Annals of Surgery, 259(1), 1–6. doi:10.1097/SLA.0b013e3182a5c8b8

Chien, J. H., Suh, I. H., Park, S. H., Mukherjee, M., Oleynikov, D., & Siu, K. C. (2013). Enhancing Fundamental Robot-Assisted Surgical Proficiency by Using a Portable Virtual Simulator. Surgical innovation, 20(2), 198-203.).

Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J. Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparos- copy. Surg Endosc. 2003;17:1082-1085.)

Green CS, Bavelier D. Action-video-game experience alters the spatial resolution of vision. Psychological Science. 2007; 18:88–94.

Green CS, Bavelier D. Enumeration versus multiple object tracking: The case of action video game players. Cognition. 2006; 101:217–245.

Giannotti, D., Patrizi, G., Di Rocco, G., Vestri, A. R., Semproni, C. P., Fiengo, L., ... & Redler, A. (2013). Play to become a surgeon: impact of Nintendo Wii training on laparoscopic skills. PloS one, 8(2), e57372

Harper JD, Kaiser S, Ebrahimi K, Lamberton GR, Hadley HR, Ruckle HC, Baldwin DD (2007) Prior video game exposure does not enhance robotic surgical performance. J Endourol 21:1207

Ju, R., Chang, P. L., Buckley, A. P., & Wang, K. C. (2012). Comparison of Nintendo Wii and PlayStation2 for enhancing laparoscopic skills. JSLS: Journal of the Society of Laparoendoscopic Surgeons, 16(4), 612.

Luca, F., Valvo, M., Ghezzi, T. L., Zuccaro, M., Cenciarelli, S., Trovato, C., … Biffi, R. (2013). Impact of robotic surgery on sexual and urinary functions after fully robotic nerve-sparing total mesorectal excision for rectal cancer. Annals of Surgery, 257(4), 672–678. doi:10.1097/SLA.0b013e318269d03b

Maan, Z. N., Gibbins, N., Al-Jabri, T., & D'Souza, A. R. (2012). The use of robotics in otolaryngology-head and neck surgery: a systematic review. American Journal of Otolaryngology, 33(1), 137–146. doi:10.1016/j.amjoto.2011.04.003

Rosenthal, R., Geuss, S., Dell-Kuster, S., Schäefer, J., Hahnloser, D., & Demartines, N. (2011). Video gaming in children improves performance on a virtual reality trainer but does not yet make a laparoscopic surgeon. Surgical innovation, 1553350610392064.

Rosser Jr, J. C., Gentile, D. A., Hanigan, K., & Danner, O. K. (2012). The effect of video game "warm-up" on performance of laparoscopic surgery tasks. JSLS: Journal of the Society of Laparoendoscopic Surgeons, 16(1), 3.

Rosser JC Jr., Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R. The impact of video games on training surgeons in the 21st century. Arch Surg. 2007;142:181–186.;

Schlickum MK, Hedman L, Enochsson L, Kjellin A, Fellander- Tsai L (2009) Systematic video game training in surgical novices improves performance in virtual reality endoscopic surgical simulators: a prospective randomized study. World J Surg 33:2360–2367

Serati, M., Bogani, G., Sorice, P., Braga, A., Torella, M., Salvatore, S., … Ghezzi, F. (2014). Robot-assisted Sacrocolpopexy for Pelvic Organ Prolapse: A Systematic Review and Meta-analysis of Comparative Studies. European Urology. doi:10.1016/j.eururo.2014.02.053

Zureikat, A. H., Moser, A. J., Boone, B. A., Bartlett, D. L., Zenati, M., & Zeh, H. J., 3rd. (2013). 250 robotic pancreatic resections: safety and feasibility. Annals of Surgery, 258(4), 554–559; discussion 559–562. doi:10.1097/SLA.0b013e3182a4e87c

Validation for Simulators: It's All About Perspective

Alyssa Tanaka, Danielle Julian, Patricia Mattingly, Roger Smith
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.Tanaka@flhosp.org, Danielle.Julian@flhosp.org,
Patricia.Mattingly@flhosp.org, Roger.Smith@flhosp.org

ABSTRACT

Does "validity" refer to the quality of an assessment, reliability of simulator outputs, or accuracy of internal simulation models? This question emerges in medical simulation and training, as educational, clinical, and engineering communities intersect. Each has developed a validation approach to meet their needs, without clear understanding of the other perspectives. Historically, validity has been assessed using a classical framework of content, criterion, and construct validity, concluding that a simulator is or is not valid. Validity has evolved into a unitary concept of construct, consisting of five distinct sources: content, response process, internal structure, relation to other variables, and consequences. Evidence for each source supports a score interpretation for a specific population, under a specific use case. This does not indicate that the assessment itself is generally valid, much less whether the simulator can be relied upon to deliver accurate results.

This unitary framework was adopted by the American Psychological Association as the standard for validating assessments and was recently endorsed as the "gold standard" for validating training tools. While this framework is effective for evaluating the appropriateness of an assessment, it may not be as robust for evaluating a *simulation device* used for assessment. This framework does not account for the physical and functional requirements of a physical system and the implications that discrepancies in those aspects may have on training and assessment.

This paper compares the classical and unitary validity methodologies with a perspective on the application to training simulators, as well as examines the inherent limitations of both. Recommendations and industry standards from other fields are also examined for applicability to surgical simulation. Finally, a recommendation for the validity classification of surgical simulators is proposed. The future of surgical certification and licensing could be reliant on simulation, however validity standards must be established to support this goal.

ABOUT THE AUTHORS

Alyssa D.S. Tanaka, Ph.D., is a Simulation Scientist at Florida Hospital Nicholson Center. Her research work focuses on robotic surgery simulation and effective surgeon training. Her current projects include rapid prototyping of surgical education devices, the validation of a robotic surgical curriculum and evaluation of robotic simulation systems. She has earned a PhD and MS in Modeling and Simulation from the University of Central Florida, as well as a Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology and Cognitive Sciences from the University of Central Florida. She holds a diploma in robotic surgery from the Department of Surgery, University of Nancy, France.

Danielle Julian, M.S., is a Research Scientist at Florida Hospital's Nicholson Center. Her current research focuses on robotic surgery simulation and effective surgeon training. . Her current projects include rapid prototyping of surgical education devices and the evaluation of robotic simulation systems. Her background includes research in Human Factors and learning and training to enhance the higher-order cognitive skills of military personnel. She is currently a Ph.D. student in Modeling and Simulation at the University of Central Florida where she previously earned an M.S. in Modeling and Simulation, Graduate Simulation Certificate in Instructional Design, and a B.S. in Psychology.

Patricia Mattingly, M.D. is a research fellow at Florida Hospital's Nicholson Center. Her research focuses on robotic surgery simulation and effective surgeon training. Her current projects include the development of a robotic surgical

curriculum and assessing surgical task-induced fatigue. She is an AAGL Fellow in Minimally Invasive Gynecologic Surgery with Columbia University Medical Center. She completed her residency in obstetrics and gynecology at Carolinas Health Care System in Charlotte, NC and earned her M.D. from the Medical College of Georgia. She received her B.S. in Biochemistry and Molecular Biology at the University of Georgia.

Roger Smith, Ph.D., is an expert in the development of simulation devices and training programs. He has spent 25 years creating leading-edge simulators for the Department of Defense and Intelligence agencies, as well as accredited methods for training with these devices. He is currently the Chief Technology Officer for the Florida Hospital Nicholson Center where he is responsible for establishing the technology strategy and leading research experiments. He has served as the CTO for the U.S. Army PEO for Simulation, Training and Instrumentation (PEO-STRI); VP and CTO for training systems at Titan Corp; and Vice President of Technology at BTG Inc. He holds a Ph.D. in Computer Science, a Doctorate in Management, and an M.S. in Statistics. He has published 3 professional textbooks on simulation, 10 book chapters, and over 100 journal and conference papers. His most recent book is *A CTO Thinks About Innovation*. He has served on the editorial boards of the *Transactions on Modeling and Computer Simulation* and the *Research Technology Management* journals.

Validation for Simulators: It's All About Perspective

Alyssa Tanaka, Danielle Julian, Patricia Mattingly, Roger Smith
Florida Hospital Nicholson Center
Celebration, FL
Alyssa.Tanaka@flhosp.org, Danielle.Julian@flhosp.org,
Patricia.Mattingly@flhosp.org, Roger.Smith@flhosp.org

INTRODUCTION

In simulation, many fields converge to create the specialized training tools used to provide learners with standardized environments for the safe acquisition of skills, relying on the expertise of engineers, educators, and subject-matter experts to create valuable training tools. It is imperative that these training systems are vetted to ensure that system performance meets the expected standards, a process typically referred to as validation. The resulting measure of validity refers to the degree to which a model or system is an accurate representation of the real world concept that it is intended to replicate (Sargent, 2000; McDougall, 2007; AERA, 1997).

The underlying validation process and associated implications are often subject to the field it is being referenced for. Using a flight simulator as an example, a computer programmer may validate the model in respect to how it performs against an actual system (e.g. aerodynamic characteristics). An engineer may assess whether the controls look and feel representative to the actual aircraft platform, and an educator validates that the flight assessment and After Action Review (AAR) accurately measure and provide relevant feedback on the trainee's performance for a specific testing context.

The surgical field has adopted virtual reality (VR) simulators, similar to flight simulators, as a solution to limited training opportunities, regulated work hours, and a need for advanced training (Kuhn, 1962; Gallagher & Sullivan, 2011). Similar to the validation of a flight simulator, each stakeholder involved in the development and implementation of a surgical simulator has a specific expectation for the concept of validity. The programmers are interested in how closely the physics models of the virtual environment are representative of the real world (e.g. how tissue behaves when retracted) and the engineers verify that the controls function similarly to the actual surgical instruments. The educators and researchers are more concerned with how the benchmarks and scoring system translate to the learners.

The introduction of VR simulators coincided with a drive in the surgical field to move away from the traditional apprenticeship model and towards proficiency-based training. This has been critically important particularly in the specialized field of robotic surgery. Currently, four VR robotic surgery simulators exist: the da Vinci Skills Simulator (dVSS) by Intuitive Surgical Inc., also known as the "Backpack Simulator"; the dV-Trainer from Mimic Technologies Inc., the RoSS by Simulated Surgical Sciences Inc., and the Robotix Mentor from Simbionix (Figure 1). While all of these systems attempt to replicate the controls, visual system, and console of the actual surgical robot, each has unique qualities in regards to software, hardware, and assessment methods.

In the dVSS, the trainee sits at and operates the simulated environment using the actual da Vinci surgeon console. The simulator is a custom computer, appended to the surgical console through the actual surgical data port. Using this simulator, users can train using the actual hardware they would use during surgery. The second is a standalone system that utilizes a graphic/gaming computer, connected to a custom desktop viewing and control device that replicates the hardware of the da Vinci surgeon's console. This system shares similar software with the dVSS, but does not require the use of any actual da Vinci hardware. The third is composed of a completely customized replica of the da Vinci surgeon's console. Internally the simulator contains a graphic computer, a 3D monitor, and commercial Omni Phantom haptic controllers (Smith, Truong, & Perez, 2014). The Robotix Mentor is a standalone system that uses custom hardware for the master controllers and Sony glasses for the 3D visual system (Robotix Mentor, n.d). These variations in hardware and software have resulted in many research studies attempting to validate these systems, as illustrated in a summary of these studies in Smith et al. (2015) and Stephanidis (2015).

Figure 1. Different aspects of surgical simulation

The validation studies that have been performed over the last decade have come at a time when medical education and assessment are shifting to new standards. Therefore, the interested educational communities have called for a shift away from the methods of previous studies and towards a new standard process. This discussion has revealed a distinct difference in the perspectives of different communities that are interested in the validation of simulators and of the educational outcomes they provide. In this paper, we present three dominant models for validation which may appear to be in conflict, but which actually represent the distinct needs of different communities, at different phases in a simulator's lifecycle. This paper also provides a process for integrating multiple validation methods for effectively assessing educational technology.

VALIDATION FRAMEWORKS

Multiple professional communities have developed validation frameworks that address their own needs to insure, measure, and certify the accuracy, realism, and assessments provided by a simulator. The work of each of these communities is just beginning to be known to members of the other communities, which is triggering both mild and vehement disagreements about the meaning, purpose, and methods of validation. Cultural and intellectual clashes of these types have occurred repeatedly in other areas of science and engineering. Those cases, as in this, are often fueled by a lack of understanding of the perspectives and needs of the conflicting communities.

In surgical simulation, several frameworks for proving validity have been proposed as the standard for validating educational technology. While the American Psychological Association (APA) endorses a "unitary" framework as the gold standard for validating assessment tools, this model alone does not account for the need to validate simulators from different perspectives in other fields. A shared understanding of all of the perspectives involved may eliminate much of the friction that is being generated in this area. The most prominent validation frameworks from three different communities is shown in Figure 2 and discussed below.

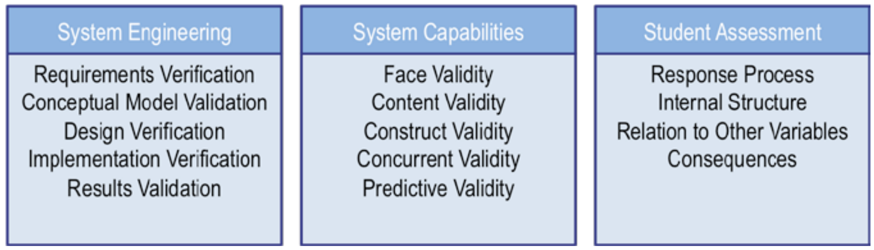


Figure 2. Summary of the validation frameworks

System Engineering Validation

The community that develops simulators and implements a formal process for validating their accuracy and usefulness has relied on Sargent's (2000) model for guidance through the engineering process, and indirectly the work of Balci (1997). In this model, the terms verification, validation, and accreditation (VV&A) are used to increase the preciseness of defining the steps in the process (Figure 3). However, this entire process is appropriately comparable to the other two frameworks that are explored in this paper.

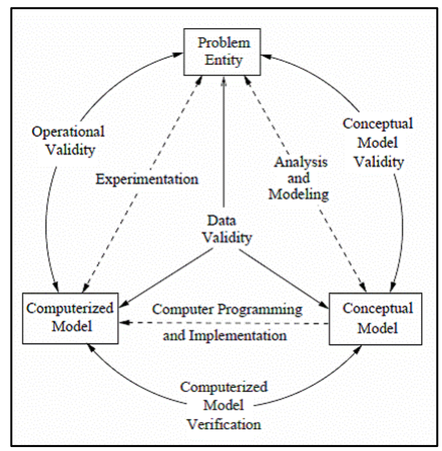


Figure 3. VV&A in Simulator Development (Sargent, 2000)

The creators and users of this framework are faced with a different set of problems than those who use of the other validation frameworks. Here, the emphasis is on guiding, controlling, modifying, and using a simulator as a hardware and software system or device. Because simulators are approximate replicas of some real world system, they can be created with dozens or hundreds of different representations of the world which may or may not be accurate and useful models of the real system and the purpose to which they are being put. This process seeks to expose the degree to which the simulator hardware, software, and data effectively represent the real world. This has to be done in the context of the expected application of the simulator. This context is essential in deciding whether compromises which have been made impact or invalidate the usefulness of the simulator in its specific application.

Sargent's framework has become the de facto validation process in the engineering and development of simulators. It is included in multiple later works which prescribe the process of simulator development and the accompanying validation of the product, such as Tolk (2012), Fishwick (2007), and others. In spite of this prevalence, the Sargent framework does not appear as a reference or an application in any of the medical simulation literature. Those communities come to simulation at a very different time in the system's lifecycle. They more typically encounter a simulator after it has been designed and manufactured for them by a device company. The users of the simulator are then more interested in the degree to which it can assist them with teaching concepts and measuring competence. So their need for validation is entirely at the user experience, educational effectiveness, and student assessment levels. In spite of the fact that the device company may have rigorously applied the VV&A methods of Sargent (2000) and Tolk (2012), the medical users will insist upon another layer of validation of the product using one of the other frameworks.

Classical Validation

To support the needs of communities using educational devices, to include simulators, the American Educational Research Association (AERA) and the American Psychological Association (APA) proposed a framework for assessing educational tools, typically referred to as the "classical" framework (AERA, 1985). The goal of this validity model is to assess educational tools to ensure that a tool is meeting the educational goals of assessing the specific abilities that it was intended to test.

Under this methodology, evidence is gathered to support a specific inference being made from test scores. For example, if a passing test score implies that a surgeon has the basic skills required to perform the removal of a prostate, then evidence would need to be gathered to support this claim. Under this framework, evidence is grouped into three categories: content related, criterion related, and construct related (Table 1).

Table 1. Summary of the Classical Framework

Validity	Meaning	Example(s)
Construct	A measure indicating the degree to which a test assesses the construct that it is intended on measuring.	What is this test supposed to measure? What is this test actually measuring?
Content	A measure of the degree to which a test's content represents a defined universe or content domain.	What is the content that needs to be tested? Is the test content representative of the actual content? Does the response type and testing format match the universe?
Criterion	A measure of the degree to which the test scores are related to one or more outcome criteria.	Can the test scores accurately predict future performance in the real world? How accurately can the test predict criterion performance?

For *construct related evidence*, information is gathered to support that the test evaluates the specific characteristics of the quality being measured (i.e. does the test evaluate what it is designed to). The construct of interest is often ingrained in the test's conceptual framework and is specific to the construct's meaning, distinguishing it from other constructs and indicating how the measure should relate to other relevant variables. Gathering evidence in this domain may also involve evaluating aspects such as test format or administration, if these circumstances affect the test meaning and interpretation.

Content evidence should demonstrate the degree to which test items, tasks, or questions are representative of a specified universe or area of content, given a proposed use of the test. Gathering evidence in this domain implies determining the content that needs to be tested and determining if the test is representative of that specific content. This also includes evaluating if the testing format and response mechanism is appropriate for the content (e.g. How is a student being assessed for a test on manual skill as opposed to critical thinking). This type of evidence often relies on expert judgment to assess the relationship between the test and the defined universe, however observation in combination with expert input is acceptable. If a test is going to be used in a way that was not originally intended, the appropriateness of original domain definition needs to be evaluated for the new use.

Criterion evidence demonstrates that test scores are systematically related to one or more relevant outcome criteria. The relationship between test scores and criterion measures may be expressed in several ways, with the goal of determining the accuracy to which the outcome criterion performance can be predicted from scores on the test. In general, there are two designs for obtaining criterion related evidence: concurrent and predictive methods. A predictive study obtains information supporting the accuracy with which test data can be used to estimate future criterion performance. A concurrent study serves the same purpose, but it obtains prediction and criterion information simultaneously.

McDougall (2007) adapted this framework for applicability to medical simulators. Under this modified framework the validation types included face, content, construct, concurrent, and predictive validity. Face validity is typically assessed informally by users and indicates whether the simulator is an accurate representation of the actual system (i.e. the realism of the simulator). Content validity is the measure of the appropriateness of the system as a teaching modality. Experts who are knowledgeable about the device typically assess this via a formal evaluation. Construct validity is the ability of a simulator to measure what it is intended to measure. Often this is characterized by the simulator's ability to differentiate between users' experience level. Concurrent validity is the extent to which the

simulator correlates with the "gold standard" for training and predictive validity is the extent to which the simulator can predict a user's future surgical performance. Collectively, concurrent and predictive validity are known as criterion validity and are used as measures of the simulator's ability to correlate trainee performance with their real life performance. Face and content validity are most effective in evaluating the ability of a simulator to train a surgeon; however construct, concurrent, and predictive validity are most useful for evaluating the effectiveness of a simulator to assess a trainee. The majority of literature surrounding the validity of surgical simulators uses these categories defined by McDougall.

Unitary Validation

The AERA and APA updated the classical framework to create a new methodology for validating educational tools, referred to as the "unitary" framework because it views validity as a unitary concept of five sources of evidence: content, response process, internal structure, relations to other variables, and consequences (Table 2). The more evidence collected, the stronger the validity argument is for the test for a specific interpretation, at any given time, for a specific population. Similar to the classic framework proposed by the AERA and APA in 1997, the assessment itself is not considered completely valid or invalid, but is more or less valid.

Table 2. Summary of the Unitary Validation

Validity	Meaning	Example(s)
Test Content	A measure of the degree to which the test's content aligns with the content domain and interpretation of scores.	Are the test items assessing the content and skills that they should?
Response Process	A measure of the degree to which the response mechanisms of the test represent the skills being tested.	Are test takers demonstrating the skills being assessed?
Internal Structure	A measure of the degree to which the format and interrelatedness of the test items aligns with the construct being measured.	Is the test organized as it should be?
Relation to Other Variables	A measure of the degree to which the scores are related to variables outside of the test.	Do the scores align with a test that is currently the gold standard?
Consequences	A measure of the potential consequences of administering the test.	Are the consequences of the test scores relevant to the test's validity?

Test content evidence refers literally to the content of the test being administered. For the purpose of this measure, "content" refers to the test items, to include the wording and formatting of the test, and procedures for administration and scoring. The evidence in this domain includes either a logical or empirical analysis of the adequacy to which the test content represents the content domain and of the relevance of the content domain to the proposed interpretation of test scores. For task-based assessments, as in the case of many simulators, test evaluators create a list of tasks required by the job via observation and advisement of a subject matter expert (SME). The SME judgment assesses the criticality and frequency related to the task performance.

Response process evidence is gathered using a theoretical or empirical analysis of the response processes of test takers, which provides evidence in respect to the appropriateness of the construct and the nature of response mechanism used by the test takers. For example, if a test assesses critical analysis and reasoning, it is important to determine whether examinees are using this skill for the given material. The evidence for this domain is typically generated from an analysis of individual responses, including feedback from test takers regarding their performance strategies or reasoning of responses. In the case of scores being generated by evaluators, evidence can be gathered from the evaluators by determining the extent to which the evaluators are consistent with the interpretation of scores.

Internal structure evidence indicates the degree to which the relationships among the test items comply with the interpretation of the test score. Evidence gathered for this domain would indicate if the items on the test support the assumptions of the inter-relatedness of the items. For example if all items on a test will form a comprehensive score, then the test items should be one-dimensional. Test items may imply several aspects of a construct being tested and evidence in this domain determines the extent to which the items' relationships align with the necessity of the test framework.

Evidence gathered in regards to the *relationship to other variables* assesses the relationship of the test score to variables that are external to the test. The external variables can include measures of criteria that the test is expected to predict and relationships to other test scores that are expected to be either convergent or discriminant (i.e. measuring the same or different constructs respectively). This evidence addresses questions about the degree to which these relationships are consistent with the construct underlying the proposed test interpretation.

Lastly, evidence regarding the *consequences* does not necessarily affect the test's validity, but helps to inform the process of assessing validity. Evidence in this domain determines if there is a consequence of administering the test and if this consequence is relevant to other domains of validity. A finding in this domain of validity is relevant to the validity of the test in general if it can be directly related to another source of validity.

SYMBIOTIC FRAMEWORKS

When applying these frameworks to a simulation system being used for education, we can see that there is not one individually that meets all requirements of a system. While assessment is an essential component of a learning experience, it is not the only aspect that a user relies on for feedback when using a simulation system. Simulators are complex devices that often rely on the replicated controls and interfaces with real-world systems, including user feedback mechanisms (e.g. haptic feedback or visual stimuli). These mechanisms enhance user experience and facilitate learning by providing formative feedback and developing user expectations on how the real-world system should perform. Some simulators, including robotic surgery simulators, provide summative feedback mechanisms to the user at the end of the simulation experience, which helps to reduce the need for a proctor during the trainings. Figure 4 provides a general example of how this information is presented to the user. This feedback is often given based on specific criteria and benchmarks that are relevant to the task that the user is performing.

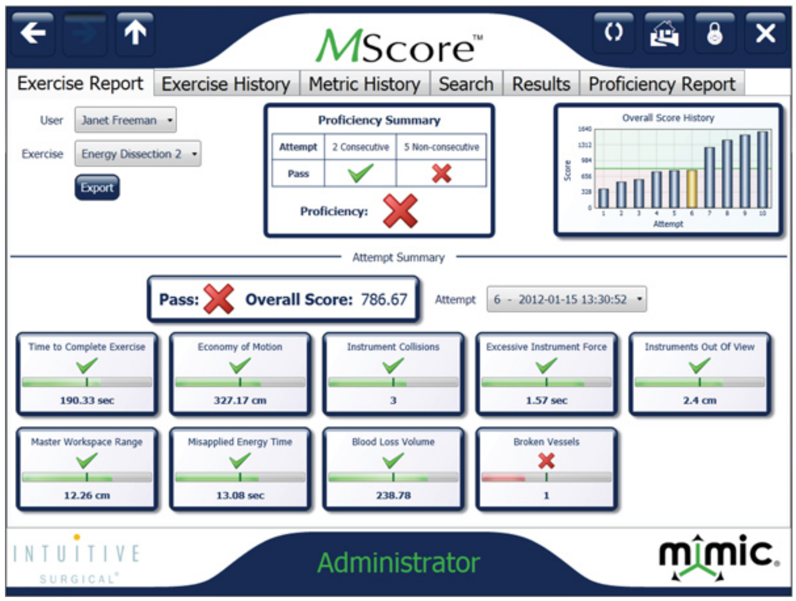


Figure 4. Robotic Surgery simulator summative feedback screen

During the simulation experience, the user makes an input into the system and receives a corresponding output from the system. For example, by moving a camera control towards a target area, the field-of-view will change to the specified location. By receiving that output the user decides what the next input will be. Using the camera example, if the user overcompensates and moves the camera past the target location, they would see this and use the camera control to adjust the field-of-view. This cycle continues until the simulation experience is complete (Figure 5).

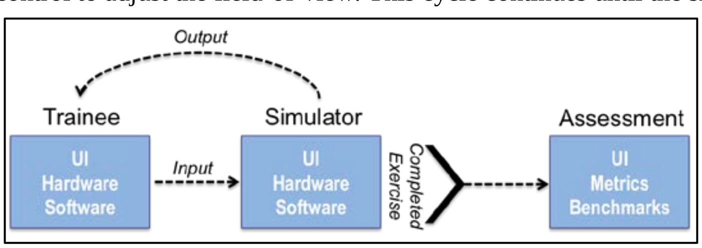
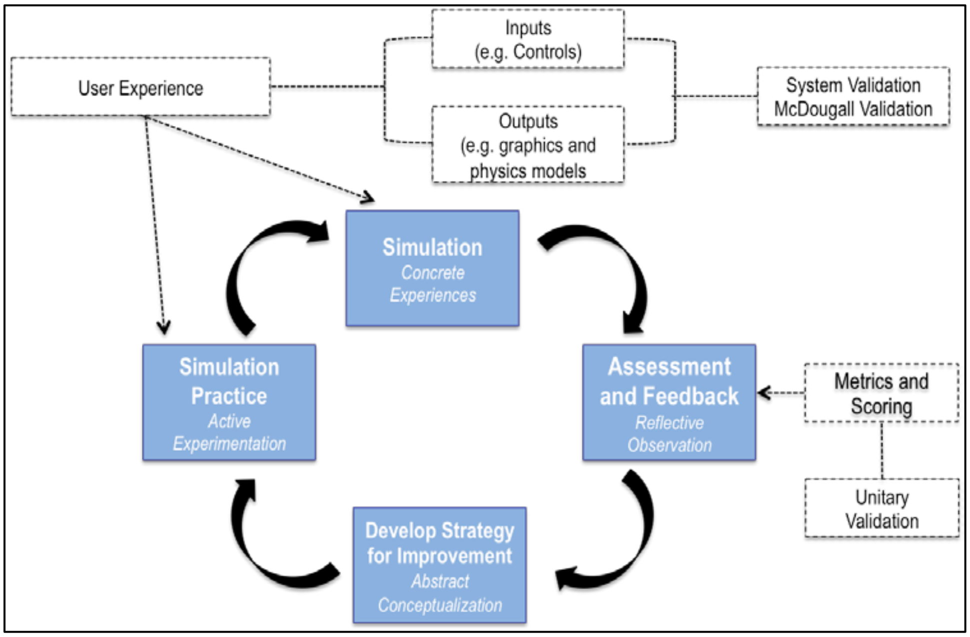


Figure 5. User interaction with simulator

The process of learning via simulation is an experiential process that can be related to the Kolb Experiential Cycle (1984) as shown in Figure 6. When looking at this model, the simulator plays a crucial role in the learning experience of the user. The user expectations are established during the *concrete experience* with the simulator. The learner applies that experience for *reflective observation* and to form an *abstract conceptualization* of how to



improve performance. Thus, the user's learning is facilitated through their interactions with the system and the formative feedback that they receive from system.

Figure 6. Image showing the relationship of the three frameworks

When looking specifically at the two educational models, the frameworks are designed for evaluating assessments and as such are focused on whether the assessment of the student was an accurate measure of the knowledge and skills that are being evaluated. If we only look at the assessment component of a simulator, then we are only looking at a small portion of the learning experience as a whole. It is possible to have a simulator that meets a high level of educational validity, but is not realistic in terms of engineering design. Conversely, we can have a simulator that almost perfectly replicates the intended system, but does not have meaningful associated metrics. In either case, the user would develop an incorrect model of their knowledge and skills during the training and assessment that would not translate to the real world system.

These frameworks cannot individually address the comprehensive needs for validation of educational simulators and thus need to be used complementarily to one another. Table 3 provides an example of different degrees of validity according to each framework which can be used to evaluate the individual simulator components and to address the needs of educators comprehensively.

Table 3. Validity Levels

	Less Validity	**Moderate Validity**	**More Validity**
Systems Engineering Framework	• Output does not match the real world measures.	• Unrealistic graphics • Pseudo-physics models.	• Highly realistic graphics • Realistic physics models.
Classical Framework (McDougall)	• Replicates real-world system to demonstrate placement of controls, but do not function the same.	• Custom hardware that is more realistic, but not exact.	• Embedded Simulator same hardware as in the real system.
Educational Framework	• Test content does not	• The content aligns	• Test content is

	align with content domain. • Test does not measure what it is intended to.	with the content domain. • The users are not demonstrating the necessary skills	relevant to the content domain. • Scores can predict future performance

CONCLUSION

This paper summarizes three prominent and valuable frameworks and demonstrates the role that each takes in the validation process. These frameworks overlap to some degree; no one framework is a complete duplication or replacement of another. Thus, the goal is to explain the rationale for the decidedly different processes that are referred to by the same term and create an awareness of these methodologies, potentially provoking adoption or adaptation. Understanding the value of different frameworks may reduce arguments and contention between communities attempting to apply their own perspective to other communities.

While valuable to specific fields, none of these validation models individually address the comprehensive needs when using simulation technologies as education and training tools. The learning experience when using a simulator encompasses components that should be evaluated distinctly to truly speak to the value of the system as an educational tool. Furthermore, disvaluing one aspect of the system during validation could have detrimental effects on the transfer of training for the user, potentially leading to negative training.

The field of simulation integrates technology, processes, and ideas from several different communities, using technology-rich learning environments to provide learners with a real-world experience for practice and assessment. To say that one method of validation alone is sufficient would be naïve. These frameworks were developed by their respective communities to address that community's specific needs, however needs of the broader simulation community require a more interdisciplinary approach.

It is imperative to critically evaluate not only about what the validation is used for, but also what the validation is evaluating and leverage the qualities of each of the validation frameworks when assessing the validity of a system. We must consider the role that each framework plays in a system and how that affects the learner.

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1985). *Standards for educational and psychological testing*. Amer Psychological Assn.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Balci, O. (1997). Principles of Simulation Model Validation, Verification, and Testing. *Transactions of the Society for Computer Simulation*. 14, 3-12.

Fishwick, P.A. editor. (2007). *Handbook of Dynamic System Modeling*. Boca Raton, FL: CRC Press.

Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.

McDougall, E.M. (2007). Validation of Surgical Simulators. *Journal of Endourology*, 21(3), 244-7.

Robotix Mentor (n.d.). Simbionix website. Retrieved from http://simbionix.com/simulators/robotix-mentor/

Sargent, R.G. (2000). Verification, Validation, and Accreditation of Simulation Models. In *Proceedings of the Winter Simulation Conference*. Piscataway, NJ: IEEE Computer Press, 106-14.

Smith, R.D., Truong, M., Perez, M. (2015). Comparative analysis of the functionality of simulators of the da Vinci surgical robot. Surgical Endoscopy, 29(4), 972-83.

Stephanidis, D. et al. (2015). Simulation in Surgery: What’s Needed Next? *Annals of Surgery*, 261(5), 846-53.

Tolk A. editor. (2012). *Engineering Principles of Combat Modeling and Distributed Simulation*. New York: John Wiley & Sons.